

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:25:44 ; Search time 44.9254 Seconds  
(without alignments)  
540.877 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	463	100.0	86	1	AAP40829	Aap40829 Sequence
2	463	100.0	86	2	AAR84061	Aar84061 Human ins
3	463	100.0	86	2	AAY42858	Aay42858 Human ins
4	463	100.0	86	3	AAB12770	Aab12770 Human pro
5	463	100.0	86	5	AAM48218	Aam48218 Human pro
6	463	100.0	86	7	ADC64463	Adc64463 Amino aci
7	463	100.0	87	1	AAP20036	Aap20036 Human pro
8	463	100.0	87	1	AAP40217	Aap40217 Sequence
9	463	100.0	87	1	AAP50127	Aap50127 Sequence

10	463	100.0	87	1	AAP50060	Aap50060	Synthetic
11	463	100.0	87	1	AAP61090	Aap61090	Sequence
12	463	100.0	87	2	AAR32367	Aar32367	Proinsuli
13	463	100.0	88	2	AAR07682	Aar07682	Modified
14	463	100.0	88	2	AAR33855	Aar33855	hpI. 3/20
15	463	100.0	92	2	AAR20467	Aar20467	Yeast alp
16	463	100.0	93	1	AAP90102	Aap90102	Synthetic
17	463	100.0	96	2	AAy08004	Aay08004	Human pro
18	463	100.0	96	2	AAO17830	Aao17830	Human pro
19	463	100.0	97	2	AAR68898	Aar68898	Human pro
20	463	100.0	97	3	AAB12773	Aab12773	Human pro
21	463	100.0	110	1	AAP10042	Aap10042	Sequence
22	463	100.0	110	1	AAP10053	Aap10053	Sequence
23	463	100.0	110	1	AAP40309	Aap40309	Sequence
24	463	100.0	110	2	AAy06608	Aay06608	Human pre
25	463	100.0	110	3	AAy44367	Aay44367	Human pro
26	463	100.0	110	3	AAy70366	Aay70366	Human Ins
27	463	100.0	110	3	AAB26765	Aab26765	Human pre
28	463	100.0	110	3	AAB06144	Aab06144	Human ins
29	463	100.0	110	4	AAE10337	Aae10337	Human pre
30	463	100.0	110	4	AAB35424	Aab35424	Secretory
31	463	100.0	110	4	AAG65677	Aag65677	Human pro
32	463	100.0	110	5	ABG60634	Abg60634	Human ins
33	463	100.0	110	5	ABG31590	Abg31590	Human pre
34	463	100.0	110	6	ABR55862	Abr55862	Humanised
35	463	100.0	110	6	ADA09218	Ada09218	Human Pre
36	463	100.0	110	7	ADC51569	Adc51569	Human pro
37	463	100.0	110	7	ADE56710	Ade56710	Human Pro
38	463	100.0	110	7	ADD46938	Add46938	Human Pro
39	463	100.0	110	7	ADE57650	Ade57650	Human Pro
40	463	100.0	117	3	AAy69788	Aay69788	MWPsp-MWP
41	463	100.0	130	3	AAy69787	Aay69787	MWPsp-MWP
42	463	100.0	151	2	AAW81856	Aaw81856	Human pro
43	463	100.0	153	3	AAy53589	Aay53589	Human pre
44	463	100.0	161	2	AAW81857	Aaw81857	Human pro
45	457	98.7	88	2	AAR39574	Aar39574	Human Met

# ALIGNMENTS

## RESULT 1

AAP40829

ID AAP40829 standard; protein; 86 AA.

XX

AC AAP40829;

XX

DT 25-MAR-2003 (revised)

DT 03-AUG-1992 (first entry)

XX

DE Sequence of human insulin precursor.

XX

KW Insulin precursor; connecting peptide; diabetes; hormone.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Region 1. .30  
 FT /label= chain B  
 FT Modified-site 1  
 FT /label= F-NH2-R  
 FT /note= "H or a chemically or enzymatically cleavable AA  
 FT residue or peptide residue"  
 FT Disulfide-bond 7. .72  
 FT Disulfide-bond 19. .85  
 FT Peptide 31. .65  
 FT /label= connecting peptide  
 FT Region 66. .86  
 FT /label= chain A  
 FT Disulfide-bond 71. .76  
 FT Modified-site 86  
 FT /label= N-OH  
 XX  
 PN US4430266-A.  
 XX  
 PD 07-FEB-1984.  
 XX  
 PF 16-FEB-1982; 82US-00349397.  
 XX  
 PR 27-MAR-1980; 80US-00134389.  
 PR 28-NOV-1980; 80US-00210696.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Frank BH;  
 XX  
 DR WPI; 1984-049032/08.  
 XX  
 PT Insulin precursor prodn. from linear S-sulphonate and mercaptan - in  
 PT single step without separate oxidn.  
 XX  
 PS Claim 17; Col 4; 8pp; English.  
 XX  
 CC The inventors claim a method for the prepn. of an insulin precursor in  
 CC which the A-chain and B-chain are joined through a connecting peptide.  
 CC The connecting peptide joins the A-chain at the amino group of A-1 to the  
 CC B-chain at the carboxyl group of B-30. The method is pref. for the prepn.  
 CC of human insulin precursor (see AAP40829). The SQs of the connecting  
 CC peptides of a number of species are given (see AAP40828, AAP40830-39).  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 1; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 2

AAR84061

ID AAR84061 standard; protein; 86 AA.

XX

AC AAR84061;

XX

DT 22-AUG-1996 (first entry)

XX

DE Human insulin.

XX

KW Insulin; transformation; gene expression; fungi; fungal cell; hormone;

KW A-chain; C-chain; glycosylation.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1. .261

FT /\*tag= a

FT /product= "Insulin."

XX

PN EP704527-A2.

XX

PD 03-APR-1996.

XX

PF 03-AUG-1995; 95EP-00112210.

XX

PR 05-AUG-1994; 94HR-00000432.

XX

PA (PLIV ) PLIVA PHARM & CHEM FAB.

XX

PI Mestric S, Punt PJ, Valinger R, Van Den Hondel CAMJJ;

XX

DR WPI; 1996-129917/18.

DR N-PSDB; AAT17830, AAT17831.

XX

PT DNA encoding human insulin precursors - which comprise B- and A-chains

PT linked via amino acid chain contg. 1 or more glycosylation sites, for

PT prepn. of insulin in fungal cells.

XX

PS Disclosure; Fig 1; 32pp; English.

XX

CC DNA sequences encoding insulin precursors of formula B-Pg-A, where B and

CC A represent B- and A-chains of insulin respectively, and Pg represents a

CC modified C-peptide or any number of amino acids comprising at least one

CC glycosylation consensus site, can be inserted into expression vectors

CC which in turn can be used to transform fungal host cells. The fungal

CC cells are then cultured and the insulin expressed in such cells can be

CC harvested

XX

SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 2; Length 86;

Best Local Similarity 100.0%; Pred. No. 1.1e-43;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 3

AA42858

ID AA42858 standard; protein; 86 AA.

XX

AC AA42858;

XX

DT 19-JAN-2000 (first entry)

XX

DE Human insulin precursor, SEQ ID 5.

XX

KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;  
 KW conformation; chimeric protein; cleavable; recombinant; production;  
 KW yield.

XX

OS Homo sapiens.

XX

PN W09950302-A1.

XX

PD 07-OCT-1999.

XX

PF 31-MAR-1998; 98WO-CN000052.

XX

PR 31-MAR-1998; 98WO-CN000052.

XX

PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX

PI Gan Z;

XX

DR WPI; 1999-610839/52.

XX

PT New chimeric proteins containing human growth hormone fragment, used  
 PT particularly for the production of human insulin.

XX

PS Claim 10; Page 29; 46pp; English.

XX

CC This sequence represents a human insulin precursor comprising insulin A  
 CC and B chains separated by a 34 residue peptide sequence. This insulin  
 CC precursor can be a component of chimeric proteins which additionally  
 CC contains an N-terminal fragment of human growth hormone (hGH) and a  
 CC cleavable peptide linker (AA42857). The hGH portion of the chimeric  
 CC protein acts as an intramolecular chaperone (IMC) for the insulin  
 CC precursor, enabling it to fold correctly. The cleavable peptide linker  
 CC has a C-terminal Arg residue which enables the hGH portion of the  
 CC chimeric protein to be removed after folding has taken place. Production  
 CC of recombinant human insulin via an hGH-proinsulin chimeric protein can  
 CC provide human insulin with correctly linked cysteine bridges with fewer  
 CC necessary procedural steps, and hence resulting in a higher yield of  
 CC human insulin. The IMC sequences not only protect insulin sequences from

CC intracellular degradation by a microorganism host, but also promote the  
CC folding of the fused insulin precursor, facilitate the solubility of the  
CC fusion protein and decrease the intermolecular interactions among the  
CC fusion proteins, thus allowing folding of the fused insulin precursor at  
CC commercially useful high concentrations. The procedural steps of cyanogen  
CC bromide cleavage, oxidative sulphytolysis and related purification steps  
CC can thus be eliminated, along with the use of high concentrations of  
CC mercaptan or the use of hydrophobic absorbent resins

XX

SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

#### RESULT 4

AAB12770

ID AAB12770 standard; protein; 86 AA.

XX

AC AAB12770;

XX

DT 22-NOV-2000 (first entry)

XX

DE Human proinsulin protein sequence SEQ ID NO:2.

XX

KW Human; insulin-like growth factor 1; IGF-1; proinsulin; insulin; mutant;  
KW variant; insulin-like growth factor binding protein; IGFBP-1; IGFBP-3;  
KW antidiabetic; neuroprotective; anorectic; tranquilliser; vulnerary;  
KW anorectic; cardiant; nephrotropic; dermatological; antiHIV; antiviral;  
KW hyperglycaemia; obesity; lung disease; glomerulonephritis;  
KW interstitial nephritis; Turner's syndrome; Laron's syndrome;  
KW short stature; increased fat mass-to-lean ratio; immunological disorder;  
KW peripheral neuropathy; multiple sclerosis; muscular dystrophy;  
KW catabolic state; trauma; wounding; infection; HIV; skin disorder;  
KW human immunodeficiency virus; diabetes; heart dysfunction;  
KW kidney disorder; whole body growth disorder.

XX

OS Homo sapiens.

XX

PN WO200040612-A1.

XX

PD 13-JUL-2000.

XX

PF 05-JAN-2000; 2000WO-US000151.

XX

PR 06-JAN-1999; 99US-0115010P.

XX

PA (GETH ) GENENTECH INC.

XX  
 PI Dubaquie Y, Lowman H;  
 XX  
 DR WPI; 2000-465955/40.  
 XX  
 PT Novel insulin-like growth factor (IGF) 1 mutants that selectively bind to  
 PT IGF binding protein (IGFBP)-1 or IGFBP-3, used to improve the half-lives  
 PT of IGF-I and insulin.  
 XX  
 PS Disclosure; Page 44; 48pp; English.  
 XX  
 CC The present invention describes an insulin-like growth factor (IGF)-1  
 CC variant (I), where an amino acid at position 3, 4, 5, 7, 10, 14, 17, 23,  
 CC 24, 25, 43, 49 or 63, optionally in combination with an amino acid at  
 CC position 12 and/or 16 of the native human IGF-1 sequence, is replaced  
 CC with an alanine, glycine, or a serine residue. The residue at position 7  
 CC may be replaced by any amino acid. (I) can have antidiabetic, cardiant,  
 CC neuroprotective, anorectic, tranquilliser, vulnerary, anorectic,  
 CC nephrotropic, dermatological, antiHIV and antiviral activities. The IGF-1  
 CC mutants are used in any methods where IGFs or insulin are used, e.g. in  
 CC treating hyperglycaemia, obesity-related, neurological, cardiac, renal,  
 CC immunological, and anabolic disorders. These disorders include lung  
 CC diseases, glomerulonephritis, interstitial nephritis, Turner's syndrome,  
 CC Laron's syndrome, short stature, increased fat mass-to-lean ratios,  
 CC immunological disorders, peripheral neuropathy, multiple sclerosis,  
 CC muscular dystrophy, catabolic states, trauma, wounding, infection, human  
 CC immunodeficiency virus (HIV), wounds, skin disorders, diabetes, heart  
 CC dysfunctions, kidney disorders, and whole body growth disorders. They can  
 CC also be used for increasing serum and tissue levels of biological active  
 CC IGF or insulin a mammal. The IGF-1 mutants improve the half-lives of IGF-  
 CC 1 and insulin. The present sequence represents the native human  
 CC proinsulin protein sequence, which is given in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 3; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 5  
 AAM48218  
 ID AAM48218 standard; protein; 86 AA.  
 XX  
 AC AAM48218;  
 XX  
 DT 18-MAR-2002 (first entry)  
 XX

DE Human proinsulin.  
 XX  
 KW Antirheumatic; antiarthritic; osteopathic; cartilage disorder;  
 KW insulin-like growth factor; IGF; binding protein; IGFBP;  
 KW rheumatoid arthritis; osteoarthritis; proinsulin; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200187323-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 16-MAY-2001; 2001WO-US015904.  
 XX  
 PR 16-MAY-2000; 2000US-0204490P.  
 PR 15-NOV-2000; 2000US-0248985P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Dubaquin Y, Filvaroff EH, Lowman HB;  
 XX  
 DR WPI; 2002-082942/11.  
 XX  
 PT Treating cartilage disorders including cartilage damage by injury or  
 PT degenerative cartilagenous disorders, by contacting cartilage with  
 PT insulin-like growth factor analog with altered affinity for IGF-binding  
 PT proteins.  
 XX  
 PS Disclosure; Fig 16; 136pp; English.  
 XX  
 CC The present invention relates to a method for treating cartilage  
 CC disorders. The method comprises contacting cartilage with an active agent  
 CC such as insulin-like growth factor (IGF-1) analog with a binding affinity  
 CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1  
 CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a  
 CC IGFBP displacer peptide that prevents the interaction of IGF with an  
 CC IGFBP and does not bind to human IGF receptor. The method is useful for  
 CC treating cartilage disorders (CD), including degenerative CD, articular  
 CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence  
 CC is human proinsulin, which was used to illustrate the invention  
 XX  
 SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 5; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 6



ADC64463

ID ADC64463 standard; protein; 86 AA.

XX

AC ADC64463;

XX

DT 18-DEC-2003 (first entry)

XX

DE Amino acid sequence for human proinsulin.

XX

KW Immunoassay; human C-peptide; HCP; immune complex; human; proinsulin.

XX

OS Homo sapiens.

XX

PN US2002160435-A1.

XX

PD 31-OCT-2002.

XX

PF 12-JUN-2001; 2001US-00878380.

XX

PR 12-JUN-2000; 2000JP-00174691.

XX

PA (KITA/) KITAJIMA S.

PA (KURA/) KURANO Y.

PA (NAKA/) NAKATSUBO K.

PA (NISH/) NISHIZONO I.

XX

PI Kitajima S, Kurano Y, Nakatsubo K, Nishizono I;

XX

DR WPI; 2003-765139/72.

XX

PT Measuring human C-peptide, by reacting sample C-peptide with two  
PT different human C-peptide antibodies that recognize different epitopes on  
PT peptide, to form immune complex, separating and quantifying immune  
PT complex.

XX

PS Disclosure; SEQ ID NO 1; 20pp; English.

XX

CC The present invention relates to an immunoassay for measuring human C-  
CC peptide (HCP). The method comprises reacting HCP in a sample with a first  
CC anti-HCP antibody and a second anti-HCP antibody which is immobilised on  
CC a support, to form an immune complex, and separating and quantifying the  
CC immune complex, where the first and second antibody recognises the  
CC epitope existing in the region from 1-110 and 1-16 amino acid residues,  
CC respectively, from the N-terminal end of HCP. Also disclosed is a kit for  
CC measuring human C-peptide. The method is useful for measuring human C-  
CC peptides. The method provides high reproducibility, high detection  
CC sensitivity, and low cross-reactivity to proinsulin. The present sequence  
CC represents the amino acid sequence for human proinsulin.

XX

SQ Sequence 86 AA;

Query Match 100.0%; Score 463; DB 7; Length 86;

Best Local Similarity 100.0%; Pred. No. 1.1e-43;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|

Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 7

AAP20036

ID AAP20036 standard; protein; 87 AA.

XX

AC AAP20036;

XX

DT 25-MAR-2003 (revised)

DT 22-JUL-1992 (first entry)

XX

DE Human proinsulin.

XX

KW Proinsulin.

XX

OS Homo sapiens.

XX

PN EP55942-A.

XX

PD 14-JUL-1982.

XX

PF 31-DEC-1981; 81EP-00306190.

XX

PR 02-JAN-1981; 81US-00222010.

PR 23-JUL-1981; 81US-00286070.

PR 02-JAN-1982; 82US-00222010.

PR 03-MAR-1982; 82US-00354287.

XX

PA (UYNY-) STATE UNIV NEW YORK.

XX

PI Inouye M, Nakamura K;

XX

DR WPI; 1982-59775E/29.

DR N-PSDB; AAN20041.

XX

PT Plasmid cloning vehicles - useful for transforming bacterial hosts to produce eukaryotic polypeptide(s).

XX

PS Disclosure; Fig 27; 114pp; English.

XX

CC The sequence comprises human proinsulin. (Updated on 25-MAR-2003 to correct PR field.)

XX

SQ Sequence 87 AA;

Query Match 100.0%; Score 463; DB 1; Length 87;

Best Local Similarity 100.0%; Pred. No. 1.1e-43;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||  
 Db 2 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 61

Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||  
 Db 62 SLQKRGIVEQCCTSICSLYQLENYCN 87

RESULT 8

AAP40217

ID AAP40217 standard; protein; 87 AA.

XX

AC AAP40217;

XX

DT 25-MAR-2003 (revised)

DT 12-FEB-1992 (first entry)

XX

DE Sequence of the 32 N-terminal AAs of proinsulin.

XX

KW Hormone; cloning vector; phage resistant.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Region	2. .31
----	--------	--------

FT		/label= B-chain
----	--	-----------------

FT	Region	32. .66
----	--------	---------

FT		/label= C-chain
----	--	-----------------

FT	Region	67. .87
----	--------	---------

FT		/label= A-chain
----	--	-----------------

XX

PN GB2126237-A.

XX

PD 21-MAR-1984.

XX

PF 01-SEP-1983; 83GB-00023468.

XX

PR 03-SEP-1982; 82US-00414290.

PR 05-SEP-1984; 84US-00647338.

XX

PA (ELIL ) LILLY & CO ELI.

XX

PI Hershberge CL, Rosteck PR;

XX

DR WPI; 1984-070793/12.

DR N-PSDB; AAN40179.

XX

PT Protecting bacteria from phage infection - by transformation with cloning  
 PT vector contg. segment with restriction and modification activity.

XX

PS Example; Fig 10; 28pp; English.

XX

CC Plasmid pTh alpha 1 was constructed by inserting a synthesised gene for  
 CC thymosin alpha 1 (AAN40178) into plasmid pBR322. It is used for the  
 CC construction of pTrp24. The inventors claim a method for protecting  
 CC bacteria from phage infection - by transformation with cloning vector  
 CC contg. segment with restriction and modification activity. Prodn. of  
 CC plasmid pPR 26 or pPR27 which uses pTrp24; and prodn. of plasmid pPR29  
 CC which uses a synthetic gene coding for the 32 N-terminal AAs of





PN EP159123-A.  
 XX  
 PD 23-OCT-1985.  
 XX  
 PF 04-MAR-1985; 85EP-00301468.  
 XX  
 PR 06-MAR-1984; 84US-00586582.  
 PR 26-JUL-1984; 84US-00634920.  
 PR 31-JAN-1985; 85US-00697090.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Hsiung HM, Schoner RG, Schoner BE;  
 XX  
 DR WPI; 1985-265090/43.  
 DR N-PSDB; AAN50082.  
 XX  
 PT New selectable and autonomously replicating DNA expression vector -  
 PT useful in producing proteinaceous granules in cell transformants, esp.  
 PT for prodn. of bovine growth hormone derivs.  
 XX  
 PS Disclosure; Fig 14; 115pp; English.  
 XX  
 CC The synthetic proinsulin gene is expressed in a new selectable and  
 CC autonomously replicating recombinant DNA expression vector comprising a  
 CC runaway replicon and a transcriptional and translational activating  
 CC sequence in the reading frame of the proinsulin coding sequence, the  
 CC sequence contg. a translational stop signal. Host cells contg. the  
 CC vector, which is esp. plasmid pCZ103, are cultured, and proinsulin is  
 CC produced as a highly homogeneous species of proteinaceous granule. The  
 CC granule can be readily isolated from cell lysates and is stable on  
 CC washing with urea or detergent solns. at low concns. The granule contains  
 CC at least 50% of proinsulin and all isolation operations are simplified.  
 CC (Updated on 16-AUG-2002 to add missing OS field.) (Updated on 25-MAR-2003  
 CC to correct PA field.)  
 XX  
 SQ Sequence 87 AA;

Query Match 100.0%; Score 463; DB 1; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 61  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||||  
 Db 62 SLQKRGIVEQCCTSICSLYQLENYCN 87

RESULT 11  
 AAP61090  
 ID AAP61090 standard; protein; 87 AA.  
 XX  
 AC AAP61090;  
 XX

DT 28-FEB-1992 (first entry)  
 XX  
 DE Sequence encoded by the structural gene for human proinsulin.  
 XX  
 KW Recombinant plasmid; E.coli expression vector; secretion vector.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US4624926-A.  
 XX  
 PD 25-NOV-1986.  
 XX  
 PF 03-MAR-1982; 82US-00354287.  
 XX  
 PR 02-JAN-1981; 81US-00222010.  
 PR 23-JUL-1981; 81US-00286070.  
 XX  
 PA (UYNY-) UNIV OF NEW YORK.  
 XX  
 PI Inouye M, Nakamura K;  
 XX  
 DR WPI; 1986-331802/50.  
 DR N-PSDB; AAN60872.  
 XX  
 PT New recombinant plasmid(s) - contg. DNA sequences encoding exogenous  
 PT polypeptide and outer membrane protein of E coli.  
 XX  
 PS Example; Fig 27; 44pp; English.  
 XX  
 CC The inventors claim new recombinant plasmids contg. a DNA sequence  
 CC encoding a polypeptide, which is foreign to E.coli, in reading phase with  
 CC a DNA SQ, coding for at least one functional fragment derived from an  
 CC outer membrane lipoprotein gene of E.coli. The foreign gene may be for  
 CC human insulin. The lipoprotein gene functional fragment may be the  
 CC promoter, the 5'-UTR, the 3'-UTR or the transcription termination signal  
 CC provided that it includes at least the promoter  
 XX  
 SQ Sequence 87 AA;

Query Match 100.0%; Score 463; DB 1; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 61  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||||||  
 Db 62 SLQKRGIVEQCCTSICSLYQLENYCN 87

RESULT 12  
 AAR32367  
 ID AAR32367 standard; protein; 87 AA.  
 XX  
 AC AAR32367;

XX  
 DT 25-MAR-2003 (revised)  
 DT 18-JUN-1993 (first entry)  
 XX  
 DE Proinsulin protein sequence.  
 XX  
 KW Human; proinsulin; vector; pUC19; pPINS; CAT; pUC-CAT-proinsulin;  
 KW insulin analogue; type I; type II; diabetes.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9303174-A1.  
 XX  
 PD 18-FEB-1993.  
 XX  
 PF 31-JUL-1992; 92WO-US006451.  
 XX  
 PR 08-AUG-1991; 91US-00741938.  
 PR 30-JUL-1992; 92US-00918953.  
 XX  
 PA (SCIO-) SCIOS INC.  
 PA (PFIZ ) PFIZER INC.  
 XX  
 PI Andy RJ, Larson ER;  
 XX  
 DR WPI; 1993-076530/09.  
 DR N-PSDB; AAQ37003.  
 XX  
 PT New hepato selective and peripheral selective human insulin analogues -  
 PT and their corresp. DNA, for treatment of type I and type II diabetes.  
 XX  
 PS Disclosure; Fig 2b; 58pp; English.  
 XX  
 CC This sequence represents human proinsulin and was decoded from the  
 CC sequences given in AAQ36996-7001. The cDNA fragment coding for proinsulin  
 CC was inserted into plasmid vector pUC19 and digested with KpnI and  
 CC HindIII. This resulted in the formation of the vector pPINS. A fragment  
 CC encoding amino acids 1-73 of CAT (see AAQ37002) was inserted into pPINS  
 CC to give a plasmid which contained DNA sequences which coded for amino  
 CC acids 1-73 of CAT, an 8 amino acid linker sequence and human proinsulin.  
 CC This plasmid, pUC-CAT-proinsulin, could be used in the formation of  
 CC insulin analogues which may be used in the treatment of types I and II  
 CC diabetes. (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 87 AA;

Query Match 100.0%; Score 463; DB 2; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |  
 Db 2 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 61  
  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |  
 Db 62 SLQKRGIVEQCCTSICSLYQLENYCN 87



RESULT 13

AAR07682

ID AAR07682 standard; protein; 88 AA.

XX

AC AAR07682;

XX

DT 25-MAR-2003 (revised)

DT 09-JAN-2003 (revised)

DT 13-FEB-1991 (first entry)

XX

DE Modified human insulin precursor.

XX

KW Human insulin precursor; cathepsin C.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Peptide	1. .2
----	---------	-------

FT		/label= N-terminal initiating dipeptide
----	--	---

FT	Peptide	3. .32
----	---------	--------

FT		/label= native human insulin B-chain
----	--	--------------------------------------

FT	Peptide	33. .67
----	---------	---------

FT		/label= natural connecting peptide of human proinsulin
----	--	--

FT	Peptide	68. .88
----	---------	---------

FT		/label= native human insulin A-chain
----	--	--------------------------------------

XX

PN EP397420-A.

XX

PD 14-NOV-1990.

XX

PF 04-MAY-1990; 90EP-00304890.

XX

PR 09-MAY-1989; 89US-00349472.

XX

PA (ELIL ) LILLY & CO ELI.

XX

PI Becker GW, Furman TC, Mackellar WC, Mcdonough JP;

XX

DR WPI; 1990-343372/46.

XX

PT Human insulin precursor - contg. Met-Tyr or Met-arg initiating di:peptide  
PT for controlled removal by cathepsin C.

XX

PS Disclosure; Page 3; 8pp; English.

XX

CC This modified human insulin precursor comprises an N-terminal initiating  
CC dipeptide, chosen from Met-Tyr or Met-Arg, which does not define a  
CC cathepsin C dipeptide removal stop point. This dipeptide is linked to the  
CC natural human insulin B-chain, natural human proinsulin connecting  
CC peptide and natural human insulin A- chain. Dipeptide removal is  
CC carefully controlled to obtain the desired prod. without further  
CC degradation occurring, irrespective of whether the next dipeptide in the  
CC sequence defines a cathepsin C stop point. (Updated on 09-JAN-2003 to add  
CC missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)

XX

SQ Sequence 88 AA;

```
Query Match      100.0%; Score 463; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.1e-43;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        |||||||
Db       3 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 62

QY     61 SLQKRGIWEQCCTSICSLYQLENYCN 86
        |||||||
Db     63 SLQKRGIWEQCCTSICSLYQLENYCN 88
```

RESULT 14

AAR33855

IP AAR33855 standard; protein; 88 AA.

XX

AC AAR33855;

XX

DT 25-MAR-2003 (revised)

DT 19-JUL-1993 (first entry)

XX

DE hpI.

XX

XX  
KW Proinsulin; hpI; native; pCZR126S; expression vector; E. coli; human;  
KW expression; immunological effect.

XX

OS Homo sapiens.

XX

PN EP534705-A2.

XX

PD 31-MAR-1993.

XX

PF 22-SEP-1992; 92EP-00308601.

XX

PR 24-SEP-1991; 91US-00764655.

XX

PA (ELIL ) LILLY & CO ELI.

XX

PI Belagaje RM;

XX

DR WPI; 1993-102806/13.

DR N-PSDB; AAQ38310.

XX

XX	
PT	Expression of low molecular wt. polypeptide(s) e.g. insulin growth factor
PT	I - by expressing as deriv. with N-terminal aminoacid to provide
PT	increased expression levels.

X2

PS Disclosure; Page 21-22; 40pp; English.

X:

XX  
CC This sequence represents an analogue of native human proinsulin (hpI).

CC This sequence represents an *in vitro* synthesized DNA encoding the  
CC The DNA encoding this sequence was used in the construction of the  
CC expression vector of the invention. The coding region of the *hpI* gene was  
CC synthesised and was cloned into the expression plasmid pCZR126S (see also  
CC AAQ38307). Expression of this gene lead to the inclusion of an extra

CC amino acid (Arg) in the second position from the N-terminal of mature  
CC hpI. The extra amino acid provides increased expression levels of the  
CC protein and is then cleaved off to avoid undesirable immunological  
CC effects when used in humans. (Updated on 25-MAR-2003 to correct PN  
CC field.)

XX

SQ Sequence 88 AA;

Query Match 100.0%; Score 463; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 3 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 62  
  
QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 63 SLQKRGIVEQCCTSICSLYQLENYCN 88

# RESULT 15

AAR20467

ID AAR20467 standard; protein; 92 AA.

XX

AC AAR20467;

XX

DT 25-MAR-2003 (revised)

DT 21-APR-1992 (first entry)

XX

DE Yeast alpha-factor signal-human proinsulin fusion product.

XX

KW BCA-5; yeast preferred codons; post-translational processing; insulin;  
KW endopeptidase.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Cleavage-site 6. .7

FT /note= "signal-proinsulin junction"

FT Cleavage-site 37. .38

XX

PN US5077204-A.

XX

PD 31-DEC-1991.

XX

PF 08-APR-1988; 88US-00183252.

XX

PR 21-JUN-1984; 84US-00623308.

XX

PA (REGC ) UNIV CALIFORNIA.

XX

PI Brake AJ, Blair LC, Julius D, Thorner JW;

XX

DR WPI; 1992-032671/04.

DR N-PSDB; AAQ20543.

XX

PT Novel DNA for endo:peptidase prodn. - useful for in vivo or in vitro  
PT processing of poly:peptide(s).  
XX  
PS Example 1; Fig 1; 16pp; English.  
XX  
CC The fusion product is encoded by a synthetic sequence having at its 5'-  
CC end a modification of the 3'-end of the naturally occurring alpha-factor  
CC secretory leader and processing signal sequence, where three Glu-Ala  
CC pairs have been deleted. A plasmid containing the synthetic proinsulin  
CC coding sequence was used to transform kex2- mutant yeast strains in the  
CC presence or absence of the cloned KEX2 gene. Post-translational  
CC processing of pro-insulin into peptides only occurred in yeast  
CC transformed to KEX2 plus. See also AAQ20545. (Updated on 25-MAR-2003 to  
CC correct PA field.)  
XX  
SQ Sequence 92 AA;

Query Match 100.0%; Score 463; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 7 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 66  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 67 SLQKRGIVEQCCTSICSLYQLENYCN 92

Search completed: July 15, 2004, 16:35:33  
Job time : 46.9254 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:30:45 ; Search time 12.9963 Seconds  
(without alignments)  
341.624 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	463	100.0	86	4	US-09-477-924-2	Sequence 2, Appli
2	463	100.0	86	4	US-09-723-981-2	Sequence 2, Appli
3	463	100.0	86	4	US-09-723-896-2	Sequence 2, Appli
4	463	100.0	86	4	US-09-878-380-1	Sequence 1, Appli
5	463	100.0	96	2	US-09-134-836-4	Sequence 4, Appli
6	463	100.0	96	4	US-09-386-303A-4	Sequence 4, Appli
7	463	100.0	97	1	US-08-160-376A-4	Sequence 4, Appli
8	463	100.0	110	3	US-08-950-720A-11	Sequence 11, Appli
9	463	100.0	110	3	US-08-589-028-2	Sequence 2, Appli
10	463	100.0	110	3	US-08-784-582-2	Sequence 2, Appli
11	463	100.0	110	3	US-08-785-271-2	Sequence 2, Appli

12	463	100.0	110	4	US-08-472-701-2	Sequence 2, Appli
13	463	100.0	110	4	US-09-185-852-2	Sequence 2, Appli
14	463	100.0	110	5	PCT-US95-08596-2	Sequence 2, Appli
15	463	100.0	117	4	US-09-280-030-63	Sequence 63, Appl
16	463	100.0	130	4	US-09-280-030-62	Sequence 62, Appl
17	463	100.0	151	2	US-08-508-664-15	Sequence 15, Appl
18	463	100.0	161	2	US-08-508-664-16	Sequence 16, Appl
19	463	100.0	167	1	US-07-918-953-8	Sequence 8, Appli
20	463	100.0	167	1	US-08-081-661-8	Sequence 8, Appli
21	457	98.7	96	2	US-09-134-836-5	Sequence 5, Appli
22	457	98.7	96	4	US-09-386-303A-5	Sequence 5, Appli
23	457	98.7	97	1	US-08-389-487-7	Sequence 7, Appli
24	456	98.5	90	1	US-08-030-731A-43	Sequence 43, Appl
25	456	98.5	98	4	US-09-701-968-7	Sequence 7, Appli
26	456	98.5	99	4	US-09-701-968-8	Sequence 8, Appli
27	456	98.5	100	4	US-09-701-968-9	Sequence 9, Appli
28	446	96.3	97	3	US-09-099-307-6	Sequence 6, Appli
29	444	95.9	97	3	US-09-099-307-8	Sequence 8, Appli
30	443	95.7	110	3	US-08-589-028-4	Sequence 4, Appli
31	443	95.7	110	3	US-08-784-582-4	Sequence 4, Appli
32	443	95.7	110	3	US-08-785-271-4	Sequence 4, Appli
33	440	95.0	97	3	US-09-099-307-7	Sequence 7, Appli
34	435	94.0	97	3	US-09-099-307-11	Sequence 11, Appl
35	398	86.0	91	4	US-09-676-787-7	Sequence 7, Appli
36	292.5	63.2	67	4	US-08-981-988A-1	Sequence 1, Appli
37	290.5	62.7	83	4	US-08-981-988A-3	Sequence 3, Appli
38	288.5	62.3	83	4	US-08-981-988A-6	Sequence 6, Appli
39	288.5	62.3	113	4	US-09-484-848-16	Sequence 16, Appl
40	286.5	61.9	67	4	US-08-981-988A-2	Sequence 2, Appli
41	285.5	61.7	83	4	US-08-981-988A-8	Sequence 8, Appli
42	284.5	61.4	67	4	US-08-981-988A-5	Sequence 5, Appli
43	281.5	60.8	67	4	US-08-981-988A-7	Sequence 7, Appli
44	281	60.7	86	4	US-09-201-227A-43	Sequence 43, Appl
45	280	60.5	112	4	US-09-484-848-15	Sequence 15, Appl

#### ALIGNMENTS

RESULT 1  
 US-09-477-924-2  
 ; Sequence 2, Application US/09477924  
 ; Patent No. 6403764  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dubaquie, Yves  
 ; APPLICANT: Lowman, Henry  
 ; TITLE OF INVENTION: PROTEIN VARIANTS  
 ; FILE REFERENCE: P1712R1-1  
 ; CURRENT APPLICATION NUMBER: US/09/477,924  
 ; CURRENT FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SEQ ID NO 2  
 ; LENGTH: 86  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-477-924-2

Query Match 100.0%; Score 463; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.8e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
```

RESULT 2

US-09-723-981-2  
; Sequence 2, Application US/09723981  
; Patent No. 6506874  
; GENERAL INFORMATION:  
; APPLICANT: Dubaquier, Yves  
; APPLICANT: Lowman, Henry  
; TITLE OF INVENTION: PROTEIN VARIANTS  
; FILE REFERENCE: P1712R1  
; CURRENT APPLICATION NUMBER: US/09/723,981  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/477,923  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-723-981-2

Query Match 100.0%; Score 463; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.8e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
```

RESULT 3

US-09-723-896-2  
; Sequence 2, Application US/09723896  
; Patent No. 6509443  
; GENERAL INFORMATION:  
; APPLICANT: Dubaquier, Yves  
; APPLICANT: Lowman, Henry  
; TITLE OF INVENTION: PROTEIN VARIANTS  
; FILE REFERENCE: P1712R1  
; CURRENT APPLICATION NUMBER: US/09/723,896  
; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: US/09/477,923  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-723-896-2

Query Match 100.0%; Score 463; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.8e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

#### RESULT 4

US-09-878-380-1

; Sequence 1, Application US/09878380  
; Patent No. 6534281  
; GENERAL INFORMATION:  
; APPLICANT: Fujirebio Inc.  
; APPLICANT: KITAJIMA, Sachiko  
; APPLICANT: KURANO, Yoshihiro  
; APPLICANT: NAKATSUBO, Kaoru  
; APPLICANT: NISHIZONO, Isao  
; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit  
Therefor

; FILE REFERENCE: 0760-0291P  
; CURRENT APPLICATION NUMBER: US/09/878,380  
; CURRENT FILING DATE: 2001-06-12  
; PRIOR APPLICATION NUMBER: JP 2000-174691  
; PRIOR FILING DATE: 2000-06-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-878-380-1

Query Match 100.0%; Score 463; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.8e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|



Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 5

US-09-134-836-4

; Sequence 4, Application US/09134836

; Patent No. 5986048

; GENERAL INFORMATION:

; APPLICANT: Rubroder, Franz-Josef

; APPLICANT: Keller, Reinhold

; TITLE OF INVENTION: Improved process for obtaining

; TITLE OF INVENTION: insulin precursors having correctly bonded cystine

bridges

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/134,836

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Leslie McDonell

; REGISTRATION NUMBER: 34,872

; REFERENCE/DOCKET NUMBER: 02481.1600-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 96 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: Protein

; LOCATION: 1..96

US-09-134-836-4

Query Match 100.0%; Score 463; DB 2; Length 96;

Best Local Similarity 100.0%; Pred. No. 2.1e-47;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

```

Db      11  FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 70
QY      61  SLQKRGIVEQCCTSICSLYQLENYCN 86
Db      71  SLQKRGIVEQCCTSICSLYQLENYCN 96

```

RESULT 6

US-09-386-303A-4

; Sequence 4, Application US/09386303A

; Patent No. 6380355

; GENERAL INFORMATION:

; APPLICANT: Rubroder, Franz-Josef

; Keller, Reinhold

; TITLE OF INVENTION: Improved process for obtaining

; insulin precursors having correctly bonded cystine

bridges

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &

; Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/386,303A

; FILING DATE: 31-Aug-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/134,836

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Leslie McDonell

; REGISTRATION NUMBER: 34,872

; REFERENCE/DOCKET NUMBER: 02481.1600-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 96 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: Protein

; LOCATION: 1..96  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-386-303A-4

Query Match 100.0%; Score 463; DB 4; Length 96;  
Best Local Similarity 100.0%; Pred. No. 2.1e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 11 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 70  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 71 SLQKRGIVEQCCTSICSLYQLENYCN 96

RESULT 7

US-08-160-376A-4

; Sequence 4, Application US/08160376A  
; Patent No. 5473049

; GENERAL INFORMATION:

; APPLICANT: Obermeier, Ranier  
; APPLICANT: Gerl, Martin  
; APPLICANT: Ludwig, Jurgen  
; APPLICANT: Sabel, Walter  
; TITLE OF INVENTION: Process For Obtaining Proinsulin  
; TITLE OF INVENTION: Possessing Correctly Linked  
; TITLE OF INVENTION: Cystine Bridges  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth A. Genoni, Esq.  
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500  
; CITY: Somerville  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08876-1258

; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM 386  
; OPERATING SYSTEM: WINDOWS 3.1  
; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/160,376A  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GE P 4240420.7  
; FILING DATE: December 2, 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Barbara V. Maurer, Esq.  
; REGISTRATION NUMBER: 31,287  
; REFERENCE/DOCKET NUMBER: HOE 92/F 384

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (908) 231-4079  
; TELEFAX: (908) 231-2255

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 97 Amino Acids  
; TYPE: Amino Acid (AA)  
; TOPOLOGY: not relevant  
US-08-160-376A-4

Query Match 100.0%; Score 463; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 2.1e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 12 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 71  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 72 SLQKRGIVEQCCTSICSLYQLENYCN 97

RESULT 8

US-08-950-720A-11

; Sequence 11, Application US/08950720A  
; Patent No. 6046028

; GENERAL INFORMATION:

; APPLICANT: Conklin, Darrell C.  
; APPLICANT: Lofton-Day, Catherine E.  
; APPLICANT: Lok, Si  
; APPLICANT: Jaspers, Stephen R.  
; TITLE OF INVENTION: INSULIN HOMOLOG  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/950,720A  
; FILING DATE:  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:  
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sawislak, Deborah A  
; REGISTRATION NUMBER: 37,438  
; REFERENCE/DOCKET NUMBER: 96-09

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-442-6672  
; TELEFAX: 206-442-6678  
; TELEX:

; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 110 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6046028e  
US-08-950-720A-11

Query Match 100.0%; Score 463; DB 3; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.5e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
  
QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 9

US-08-589-028-2

; Sequence 2, Application US/08589028  
; Patent No. 6087129

; GENERAL INFORMATION:

; APPLICANT: Newgard, Christopher B.  
; APPLICANT: Halban, Philippe  
; APPLICANT: No. 6087129mington, Karl D.  
; APPLICANT: Clark, Samuel A.  
; APPLICANT: Thigpen, Anice E.  
; APPLICANT: Quaade, Christian  
; APPLICANT: Kruse, Fred

; TITLE OF INVENTION: Recombinant Expression of Proteins From  
; TITLE OF INVENTION: Secretory Cell Lines  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/589,028  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Highlander, Steven L.  
; REGISTRATION NUMBER: 47,642  
; REFERENCE/DOCKET NUMBER: UTSD:426\HYL

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-589-028-2
```

```
Query Match          100.0%; Score 463; DB 3; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.5e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 10

US-08-784-582-2

```
; Sequence 2, Application US/08784582
; Patent No. 6110707
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Halban, Philippe A.
; APPLICANT: No. 6110707mington, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quaade, Christian
; APPLICANT: Kruse, Fred
; APPLICANT: McGarry, Dennis
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
; TITLE OF INVENTION: SECRETORY CELL LINES
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
```

```
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,582
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 60/028,427
; FILING DATE: 15-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/589,028
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-784-582-2
```

```
Query Match          100.0%; Score 463; DB 3; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.5e-47;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          ||||||||||||||||||||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 11

US-08-785-271-2

```
; Sequence 2, Application US/08785271
; Patent No. 6194176
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Halban, Philippe A.
; APPLICANT: No. 6194176mington, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quaade, Christian
; APPLICANT: Kruse, Fred
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
; TITLE OF INVENTION: SECRETORY CELL LINES
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/785,271
;     FILING DATE:  Concurrently Herewith
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/589,028
;     FILING DATE:  19-JAN-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Highlander, Steven L.
;     REGISTRATION NUMBER:  37,642
;     REFERENCE/DOCKET NUMBER:  UTSD:513
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  512/418-3000
;     TELEFAX:  512/474-7577
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  110 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-785-271-2

```

```

Query Match          100.0%;  Score 463;  DB 3;  Length 110;
Best Local Similarity 100.0%;  Pred. No. 2.5e-47;
Matches   86;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

```

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
        ||||||||||||||||||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110

```

RESULT 12

US-08-472-701-2

```

; Sequence 2, Application US/08472701
; Patent No. 6509165
; GENERAL INFORMATION:
;   APPLICANT:  Griffin, Ann C.
;   APPLICANT:  Hickey, William F.
;   TITLE OF INVENTION:  Detection and Treatment Methods for
;   TITLE OF INVENTION:  Type I Diabetes
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  LAHIVE & COCKFIELD
;     STREET:  60 State Street, suite 510
;     CITY:  Boston
;     STATE:  Massachusetts
;     COUNTRY:  USA
;     ZIP:  02109-1875
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk

```



```

;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/472,701
;   FILING DATE:
;   CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US 08/272,220
;   FILING DATE:  08-JULY-1994
;   CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  DeConti, Giulio A., Jr.
;   REGISTRATION NUMBER:  31,503
;   REFERENCE/DOCKET NUMBER:  DCI-092DV
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (617)227-7400
;   TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  110 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-472-701-2

```

```

Query Match          100.0%;  Score 463;  DB 4;  Length 110;
Best Local Similarity 100.0%;  Pred. No. 2.5e-47;
Matches   86;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

```

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
        ||||||||||||||||||||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110

```

# RESULT 13

US-09-185-852-2

```

; Sequence 2, Application US/09185852
; Patent No. 6537806
; GENERAL INFORMATION:
; APPLICANT: Osborne, William R.A.
; APPLICANT: Ramesh, Nagarajan
; TITLE OF INVENTION: Compositions and Methods for Treating Diabetes
; FILE REFERENCE: P-UW 3264
; CURRENT APPLICATION NUMBER: US/09/185,852
; CURRENT FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/087,660
; EARLIER FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 110
; TYPE: PRT

```

; ORGANISM: Homo sapiens  
US-09-185-852-2

Query Match 100.0%; Score 463; DB 4; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.5e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 14

PCT-US95-08596-2

; Sequence 2, Application PC/TUS9508596

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting

; TITLE OF INVENTION: and Treating Type I Diabetes

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 State Street, suite 510

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/08596

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/272,220

; FILING DATE: 08-JULY-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: DeConti, Giulio A., Jr.

; REGISTRATION NUMBER: 31,503

; REFERENCE/DOCKET NUMBER: DCI-092PC

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 110 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US95-08596-2

Query Match 100.0%; Score 463; DB 5; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.5e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 15

US-09-280-030-63

; Sequence 63, Application US/09280030A

; Patent No. 6506595

; GENERAL INFORMATION:

; APPLICANT: Sato, Seiji

; APPLICANT: Higashikuni, Naohiko

; APPLICANT: Kudo, Toshiyuki

; APPLICANT: Kondo, Masaaki

; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR

; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE

; TITLE OF INVENTION: DNAS

; FILE REFERENCE: 382.1026

; CURRENT APPLICATION NUMBER: US/09/280,030A

; CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: JP10-87339/1998

; EARLIER FILING DATE: 1998-03-31

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 63

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Designated is

; OTHER INFORMATION: an amino acid sequence of

; OTHER INFORMATION: MWPsp-MWPmpl0-Met-Proinsulin

US-09-280-030-63

Query Match 100.0%; Score 463; DB 4; Length 117;  
Best Local Similarity 100.0%; Pred. No. 2.7e-47;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      32 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 91

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      92 SLQKRGIVEQCCTSICSLYQLENYCN 117
```

Search completed: July 15, 2004, 16:42:31  
Job time : 13.9963 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:19 ; Search time 9.62687 Seconds  
(without alignments)  
859.311 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	463	100.0	110	1	IPHU	insulin precursor
2	463	100.0	110	2	A42179	insulin precursor
3	456	98.5	110	2	B42179	insulin precursor
4	456	98.5	110	2	JQ0178	insulin precursor
5	424	91.6	110	1	INRB	insulin precursor
6	417	90.1	110	1	IPDG	insulin precursor
7	394	85.1	86	1	IPHO	insulin precursor
8	394	85.1	110	1	IPRT2	insulin 2 precursor
9	394	85.1	110	1	INMS2	insulin 2 precursor
10	392	84.7	108	2	A39883	insulin precursor
11	392	84.7	110	2	I48166	insulin precursor
12	385	83.2	110	1	IPRT1	insulin 1 precursor
13	383	82.7	84	1	IPPG	insulin precursor

14	366.5	79.2	105	1	IPBO	insulin precursor
15	366	79.0	108	1	INMS1	insulin 1 precurs
16	334.5	72.2	108	2	S09278	insulin precursor
17	320.5	69.2	77	1	INSH	insulin precursor
18	314	67.8	110	1	IPGP	insulin precursor
19	277.5	59.9	109	1	IPRTDU	insulin precursor
20	276.5	59.7	103	2	I51221	insulin precursor
21	265.5	57.3	106	1	IPXL2	insulin II precurs
22	265.5	57.3	107	1	IPCH	insulin precursor
23	262.5	56.7	106	1	IPXL1	insulin I precurs
24	256.5	55.4	51	1	INWHP	insulin - sperm wh
25	256.5	55.4	51	1	INWHF	insulin - finback
26	256.5	55.4	51	1	INEL	insulin - elephant
27	256.5	55.4	81	1	IPDK	insulin precursor
28	256	55.3	96	2	PC7082	epidermal growth f
29	254.5	55.0	51	1	INHY	insulin - hamster
30	251.5	54.3	51	1	INMSSP	insulin - Egyptian
31	250.5	54.1	51	2	A59151	insulin precursor
32	246.5	53.2	51	1	INWH1S	insulin - sei whal
33	246.5	53.2	51	1	INGT	insulin - goat
34	246.5	53.2	51	1	INCMA	insulin - Arabian
35	245.5	53.0	51	1	INCT	insulin - cat
36	244.5	52.8	51	1	INMKSQ	insulin - common s
37	239.5	51.7	51	2	JQ0362	insulin - North Am
38	234.5	50.6	51	1	INCB	insulin - Chinchil
39	231.5	50.0	51	1	INGS	insulin - goose
40	227.5	49.1	51	1	INPQ	insulin - crested
41	227.5	49.1	51	1	INTK	insulin - turkey (
42	227.5	49.1	51	1	INOS	insulin - ostrich
43	227.5	49.1	51	1	A61129	insulin - black-be
44	227.5	49.1	51	2	A60414	insulin - slider t
45	225	48.6	52	2	S44469	insulin I1 - North

# ALIGNMENTS

## RESULT 1

IPHU

insulin precursor [validated] - human

N;Alternate names: preproinsulin

C;Species: Homo sapiens (man)

C;Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 08-Dec-2000

C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58661

R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A;Title: Sequence of the human insulin gene.

A;Reference number: A93222; MUID:80120725; PMID:6243748

A;Accession: A93222

A;Molecule type: DNA

A;Residues: 1-110 <BEL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.

Science 209, 612-615, 1980

A;Title: Genetic variation in the human insulin gene.

A;Reference number: A94253; MUID:80236313; PMID:6248962  
 A;Accession: A94253  
 A;Molecule type: DNA  
 A;Residues: 1-110 <ULL>  
 A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828  
 R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.  
 Nature 282, 525-527, 1979  
 A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.  
 A;Reference number: A93216; MUID:80054779; PMID:503234  
 A;Accession: A93216  
 A;Molecule type: mRNA  
 A;Residues: 1-110 <BEL2>  
 A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828  
 R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.  
 Science 208, 57-59, 1980  
 A;Title: Nucleotide sequence of human preproinsulin complementary DNA.  
 A;Reference number: A94251; MUID:80147417; PMID:6927840  
 A;Accession: A94251  
 A;Molecule type: mRNA  
 A;Residues: 1-110 <SUR>  
 A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828  
 R;Nicol, D.S.H.W.; Smith, L.F.  
 Nature 187, 483-485, 1960  
 A;Title: Amino-acid sequence of human insulin.  
 A;Reference number: A93144  
 A;Accession: A93144  
 A;Molecule type: protein  
 A;Residues: 25-54;90-110 <NIC>  
 R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.  
 J. Biol. Chem. 246, 1375-1386, 1971  
 A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pancreatic C-peptide.  
 A;Reference number: A92075; MUID:71116410; PMID:5101771  
 A;Accession: A92075  
 A;Molecule type: protein  
 A;Residues: 57-87 <OYE>  
 R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.  
 Eur. J. Biochem. 20, 190-199, 1971  
 A;Title: Amino acid sequence of the C-peptide of human proinsulin.  
 A;Reference number: A91186; MUID:71257722; PMID:5560404  
 A;Accession: A91186  
 A;Molecule type: protein  
 A;Residues: 57-87 <KOA>  
 R;Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell, J.I.  
 Nature Genet. 4, 305-310, 1993  
 A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR.  
 A;Reference number: I58114; MUID:93364428; PMID:8358440  
 A;Accession: I58114  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-59,63-110 <RES>  
 A;Cross-references: GB:L15440; NID:g307071; PIDN:AAA59179.1; PID:g307072  
 R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.  
 Helv. Chim. Acta 57, 2617-2621, 1974

A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.  
 A;Reference number: A91636; MUID:75077277; PMID:4443293  
 A;Contents: annotation; synthesis  
 A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical with the natural hormone in chemical and biological activities  
 A;Note: article in German with English abstract  
 R;Naithani, V.K.  
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973  
 A;Title: The synthesis of C-peptide of human proinsulin.  
 A;Reference number: A91658; MUID:75040007; PMID:4803504  
 A;Contents: annotation; synthesis of residues 57-87  
 R;Geiger, R.; Jaeger, G.; Koenig, W.  
 Chem. Ber. 106, 2347-2352, 1973  
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9,Gln-11] analogue.  
 A;Reference number: A90914  
 A;Contents: annotation; synthesis of residues 57-87  
 R;Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.  
 Biochem. J. 310, 869-874, 1995  
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction.  
 A;Reference number: S58661; MUID:96013185; PMID:7575420  
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing  
 C;Genetics:  
 A;Gene: GDB:INS  
 A;Cross-references: GDB:119349; OMIM:176730  
 A;Map position: 11p15.5-11p15.5  
 A;Introns: 63/1  
 C;Superfamily: insulin  
 C;Keywords: hormone; pancreas  
 F;1-24/Domain: signal sequence #status predicted <SIG>  
 F;25-54/Domain: insulin chain B #status experimental <BCH>  
 F;25-54,90-110/Product: insulin #status experimental <MAT>  
 F;57-87/Domain: connecting C peptide #status experimental <CPEP>  
 F;90-110/Domain: insulin chain A #status experimental <ACH>  
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 100.0%; Score 463; DB 1; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110

```

# RESULT 2

A42179

insulin precursor - chimpanzee

C;Species: Pan troglodytes (chimpanzee)

C;Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999



C;Accession: A42179; S22058  
 R;Seino, S.; Bell, G.I.; Li, W.H.  
 Mol. Biol. Evol. 9, 193-203, 1992  
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.  
 A;Reference number: A42179; MUID:92219953; PMID:1560757  
 A;Accession: A42179  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-110 <SEI>  
 A;Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252  
 A;Note: sequence extracted from NCBI backbone (NCBIP:95067)  
 C;Genetics:  
 A;Introns: 63/1  
 C;Superfamily: insulin

Query Match 100.0%; Score 463; DB 2; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-43;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

# RESULT 3

B42179

insulin precursor - green monkey

C;Species: Cercopithecus aethiops (green monkey, grivet)

C;Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999

C;Accession: B42179; A05232; S16494; S22056

R;Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.

A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: B42179

A;Molecule type: DNA

A;Residues: 1-110 <SEI>

A;Cross-references: EMBL:X61092; NID:g22808; PIDN:CAA43405.1; PID:g22809

A;Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)

R;Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin C-peptides by a semi-micro Edman degradation procedure.

A;Reference number: A92111; MUID:72258016; PMID:4626369

A;Accession: A05232

A;Molecule type: protein

A;Residues: 57-87 <PET>

C;Genetics:

A;Introns: 63/1

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-54/Domain: insulin chain B #status predicted <BCH>  
F;25-54,90-110/Product: insulin #status predicted <MAT>  
F;57-87/Domain: connecting peptide #status experimental <CPEP>  
F;90-110/Domain: insulin chain A #status predicted <ACH>  
F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 98.5%; Score 456; DB 2; Length 110;  
Best Local Similarity 98.8%; Pred. No. 1.7e-42;  
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

#### RESULT 4

JQ0178

insulin precursor - crab-eating macaque

C;Species: *Macaca fascicularis* (crab-eating macaque)

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 16-Jul-1999

C;Accession: JQ0178

R;Wetekom, W.; Groneberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.  
Gene 19, 179-183, 1982

A;Title: The nucleotide sequence of cDNA coding for preproinsulin from the  
primate *Macaca fascicularis*.

A;Reference number: JQ0178; MUID:83080474; PMID:6184262

A;Accession: JQ0178

A;Molecule type: mRNA

A;Residues: 1-110 <WET>

A;Cross-references: GB:J00336; NID:g342121; PIDN:AAA36849.1; PID:g342122

C;Superfamily: insulin

F;1-24/Domain: signal sequence #status predicted <SIG>

F;25-54,90-110/Product: insulin #status predicted <MAT>

F;25-54/Domain: insulin chain B #status predicted <BCH>

F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>

F;90-110/Domain: insulin chain A #status predicted <ACH>

F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 98.5%; Score 456; DB 2; Length 110;  
Best Local Similarity 98.8%; Pred. No. 1.7e-42;  
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

#### RESULT 5

INRB

insulin precursor - rabbit

N;Alternate names: preproinsulin

C;Species: *Oryctolagus cuniculus* (domestic rabbit)

C;Date: 24-Apr-1984 #sequence\_revision 23-Aug-1997 #text\_change 18-Jun-1999

C;Accession: A53438; A01581

R;Devaskar, S.U.; Giddings, S.J.; Rajakumar, P.A.; Carnaghi, L.R.; Menon, R.K.; Zahm, D.S.

J. Biol. Chem. 269, 8445-8454, 1994

A;Title: Insulin gene expression and insulin synthesis in mammalian neuronal cells.

A;Reference number: A53438; MUID:94179230; PMID:8132571

A;Accession: A53438

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-110 <DEV>

A;Cross-references: GB:U03610; NID:g467970; PIDN:AAA19033.1; PID:g467971

R;Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A;Title: Species variation in the amino acid sequence of insulin.

A;Reference number: A90029; MUID:66160119; PMID:5949593

A;Accession: A01581

A;Molecule type: protein

A;Residues: 25-54;90-110 <SMI>

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-24/Domain: signal sequence #status predicted <SIG>

F;25-54/Domain: insulin chain B #status experimental <BCH>

F;25-54,90-110/Product: insulin #status experimental <MAT>

F;57-87/Domain: connecting C peptide #status predicted <CPEP>

F;90-110/Domain: insulin chain A #status experimental <ACH>

F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 91.6%; Score 424; DB 1; Length 110;

Best Local Similarity 90.7%; Pred. No. 5.1e-39;

Matches 78; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|||||:|||||

Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSRREVEELQVGQAEELGGGPGAGGLQPSALEL 84

Qy 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
:|||||

Db 85 ALQKRGIVEQCCTSIQSLYQLENYCN 110

RESULT 6

IPDG

insulin precursor - dog

C;Species: *Canis lupus familiaris* (dog)

C;Date: 24-Apr-1984 #sequence\_revision 15-Nov-1984 #text\_change 16-Jul-1999

C;Accession: A92413; A01587; S16493

R;Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.

J. Biol. Chem. 258, 2357-2363, 1983

A;Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded amino acid sequence of canine preproinsulin predicts an additional C-peptide fragment.





A;Reference number: A92120; MUID:73061498; PMID:4640931  
 A;Accession: C92120  
 A;Molecule type: protein  
 A;Residues: 57-87 <TAG>  
 R;Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.  
 Ann. N. Y. Acad. Sci. 343, 425-432, 1980  
 A;Title: The structure of rat preproinsulin genes.  
 A;Reference number: I51945; MUID:80240379; PMID:6249167  
 A;Accession: I64880  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-110 <RES>  
 A;Cross-references: GB:M25585; NID:g204950; PIDN:AAA41440.1; PID:g204952  
 C;Genetics:  
 A;Gene: INS2  
 A;Introns: 63/1  
 C;Superfamily: insulin  
 C;Keywords: hormone; pancreas  
 F;1-24/Domain: signal sequence #status predicted <SIG>  
 F;25-54/Domain: insulin chain B #status experimental <BCH>  
 F;25-54,90-110/Product: insulin #status experimental <MAT>  
 F;57-87/Domain: connecting peptide #status experimental <CPEP>  
 F;90-110/Domain: insulin chain A #status experimental <ACH>  
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 85.1%; Score 394; DB 1; Length 110;  
 Best Local Similarity 84.9%; Pred. No. 9.5e-36;  
 Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        || ||||| :||| || || :||| || || || ||
Db      25 FVKQHLCGSHLVEALYLVCGERGFFYTPMSRREVEDPQVAQLLEGGGPGAGDLQTLALEV 84

Qy      61 SLQKRGIVEQCCTSI CSLYQLENYCN 86
        : ||||| :||| ||||| |||||
Db      85 ARQKRGIVDQCCTSI CSLYQLENYCN 110
  
```

#### RESULT 9

##### INMS2

insulin 2 precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 31-Mar-1992 #sequence\_revision 14-Jul-1994 #text\_change 18-Jun-1999

C;Accession: A26342; B48172; A61012; B01592

R;Wentworth, B.M.; Schaefer, I.M.; Villa-Komaroff, L.; Chirgwin, J.M.  
 J. Mol. Evol. 23, 305-312, 1986

A;Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.

A;Reference number: A92965; MUID:87169768; PMID:3104603

A;Accession: A26342

A;Molecule type: DNA

A;Residues: 1-110 <WEN>

A;Cross-references: GB:X04724; NID:g52714; PIDN:CAA28433.1; PID:g52715

R;Sawa, T.; Ohgaku, S.; Morioka, H.; Yano, S.

J. Mol. Endocrinol. 5, 61-67, 1990

A;Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the NON mouse, an animal model of human non-obese, non-insulin-dependent diabetes mellitus.

A;Reference number: A48172; MUID:90372989; PMID:2397023

A;Accession: B48172

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-110 <SAW>

R;Linde, S.; Nielsen, J.H.; Hansen, B.; Welinder, B.S.

J. Chromatogr. 462, 243-254, 1989

A;Title: Reversed-phase high-performance liquid chromatographic analyses of insulin biosynthesis in isolated rat and mouse islets.

A;Reference number: A61012; MUID:89292078; PMID:2661585

A;Accession: A61012

A;Molecule type: protein

A;Residues: 57-87 <LIN>

R;Buenzli, H.F.; Glatthaar, B.; Kunz, P.; Muelhaupt, E.; Humbel, R.E.

Hoppe-Seyler's Z. Physiol. Chem. 353, 451-458, 1972

A;Title: Amino acid sequence of the two insulins from mouse (*Mus musculus*).

A;Reference number: A01592; MUID:72189455; PMID:5063718

A;Accession: B01592

A;Molecule type: protein

A;Residues: 25-54;90-110 <BUE>

C;Genetics:

A;Introns: 63/1

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-24/Domain: signal sequence #status predicted <SIG>

F;25-54/Domain: insulin chain B #status experimental <BCH>

F;25-54,90-110/Product: insulin #status experimental <MAT>

F;57-87/Domain: connecting peptide #status experimental <CPEP>

F;90-110/Domain: insulin chain A #status experimental <ACH>

F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 85.1%; Score 394; DB 1; Length 110;

Best Local Similarity 84.9%; Pred. No. 9.5e-36;

Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

|| ||||| :||| || || :||| || || |||

Db 25 FVKQHLCGSHLVEALYLVCGERGFFYTPMSRREVEDPQVAQLELGGGPGAGDLQTLALEV 84

Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

: ||||| :||| ||||| |||||

Db 85 AQQKRGIVDQCCTSICSLYQLENYCN 110

RESULT 10

A39883

insulin precursor - douroucouli

C;Species: *Aotus trivirgatus* (douroucouli, night monkey, owl monkey)

C;Date: 27-Nov-1991 #sequence\_revision 27-Nov-1991 #text\_change 16-Jul-1999

C;Accession: A39883

R;Seino, S.; Steiner, D.F.; Bell, G.I.

Proc. Natl. Acad. Sci. U.S.A. 84, 7423-7427, 1987

A;Title: Sequence of a New World primate insulin having low biological potency and immunoreactivity.

A;Reference number: A39883; MUID:88041119; PMID:3118367  
A;Accession: A39883  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-108 <SEI>  
A;Cross-references: GB:J02989; NID:g176555; PIDN:AAA35374.1; PID:g176556  
C;Superfamily: insulin

Query Match 84.7%; Score 392; DB 2; Length 108;  
Best Local Similarity 84.9%; Pred. No. 1.5e-35;  
Matches 73; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      25 FVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDLQVGQVELGGGSITGSLPP--LEG 82

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          :|||:|:|||||:|||||
Db      83 PMQKRGVVDQCCTSICSLYQLQNYCN 108
```

RESULT 11

I48166

insulin precursor - golden hamster

C;Species: Mesocricetus auratus (golden hamster)

C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 16-Jul-1999

C;Accession: I48166

R;Bell, G.I.; Sanchez-Pescador, R.

Diabetes 33, 297-300, 1984

A;Title: Sequence of a cDNA encoding Syrian hamster preproinsulin.

A;Reference number: I48166; MUID:84133036; PMID:6365663

A;Accession: I48166

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-110 <RES>

A;Cross-references: GB:M26328; NID:g191420; PIDN:AAA37089.1; PID:g305360

C;Superfamily: insulin

Query Match 84.7%; Score 392; DB 2; Length 110;  
Best Local Similarity 84.9%; Pred. No. 1.6e-35;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSRRGVEDPQVAQLELGGGPGADDLQTLALEV 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          : |||||:|||||:|||||
Db      85 AQQKRGIVDQCCTSICSLYQLENYCN 110
```

RESULT 12

IPRT1

insulin 1 precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 23-Oct-1981 #sequence\_revision 23-Oct-1981 #text\_change 24-Sep-1999

C;Accession: A90788; A90789; A94231; B92120; I51945; A01589



R;Cordell, B.; Bell, G.; Tischer, E.; DeNoto, F.M.; Ullrich, A.; Pictet, R.; Rutter, W.J.; Goodman, H.M.  
 Cell 18, 533-543, 1979  
 A;Title: Isolation and characterization of a cloned rat insulin gene.  
 A;Reference number: A90788; MUID:80045034; PMID:498283  
 A;Accession: A90788  
 A;Molecule type: DNA  
 A;Residues: 1-110 <COR>  
 A;Cross-references: GB:J00747; NID:g204956; PIDN:AAA41442.1; PID:g204957  
 R;Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tizard, R.  
 Cell 18, 545-558, 1979  
 A;Title: The structure and evolution of the two nonallelic rat preproinsulin genes.  
 A;Reference number: A90789; MUID:80045035; PMID:498284  
 A;Accession: A90789  
 A;Molecule type: DNA  
 A;Residues: 1-110 <LOM>  
 A;Cross-references: GB:J00747; NID:g204956; PIDN:AAA41442.1; PID:g204957  
 R;Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aten, B.; Oyer, P.E.  
 Recent Prog. Horm. Res. 25, 207-282, 1969  
 A;Title: Proinsulin and the biosynthesis of insulin.  
 A;Reference number: A94231; MUID:70067613; PMID:4311938  
 A;Accession: A94231  
 A;Molecule type: protein  
 A;Residues: 25-54;90-110 <STE>  
 R;Tager, H.S.; Steiner, D.F.  
 J. Biol. Chem. 247, 7936-7940, 1972  
 A;Title: Primary structures of the proinsulin connecting peptides of the rat and horse.  
 A;Reference number: A92120; MUID:73061498; PMID:4640931  
 A;Accession: B92120  
 A;Molecule type: protein  
 A;Residues: 57-87 <TAG>  
 R;Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.  
 Ann. N. Y. Acad. Sci. 343, 425-432, 1980  
 A;Title: The structure of rat preproinsulin genes.  
 A;Reference number: I51945; MUID:80240379; PMID:6249167  
 A;Accession: I51945  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-110 <RES>  
 A;Cross-references: GB:M25584; NID:g204947; PIDN:AAA41439.1; PID:g204948  
 C;Genetics:  
 A;Gene: INS1  
 C;Superfamily: insulin  
 C;Keywords: hormone; pancreas  
 F;1-24/Domain: signal sequence #status predicted <SIG>  
 F;25-54/Domain: insulin chain B #status experimental <BCH>  
 F;25-54,90-110/Product: insulin #status experimental <MAT>  
 F;57-87/Domain: connecting peptide #status experimental <CPEP>  
 F;90-110/Domain: insulin chain A #status experimental <ACH>  
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 83.2%; Score 385; DB 1; Length 110;  
 Best Local Similarity 83.7%; Pred. No. 9e-35;

Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 || ||||| ||||||||||||||||||||:|||| || || |:||||| || || |||||  
 Db 25 FVKQHLCGPHLVEALYLVCGERGFFYTPKSRREVEDPQVPQLELGGGPEAGDLQTLALEV 84

Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 : ||||||:|||||||||||||||||||  
 Db 85 ARQKRGIVDQCCTSICSLYQLENYCN 110

RESULT 13

IPPG

insulin precursor - pig

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 22-Jun-1981 #sequence\_revision 22-Jun-1981 #text\_change 16-Jul-1999

C;Accession: A01583; A94572; S16492; A60835; B60835

R;Chance, R.E.; Ellis, R.M.; Bromer, W.W.

Science 161, 165-167, 1968

A;Title: Porcine proinsulin: characterization and amino acid sequence.

A;Reference number: A94240; MUID:68286485; PMID:5657063

A;Accession: A01583

A;Molecule type: protein

A;Residues: 1-34,'Q',36-84 <CHA>

R;Chance, R.E.

submitted to the Atlas, July 1970

A;Reference number: A94572

A;Accession: A94572

A;Molecule type: protein

A;Residues: 1-84 <CH2>

R;Brown, H.; Sanger, F.; Kitai, R.

Biochem. J. 60, 556-565, 1955

A;Title: The structure of pig and sheep insulins.

A;Reference number: A90344

A;Accession: S16492

A;Molecule type: protein

A;Residues: 1-30;31-51 <BRO>

R;Snel, L.; Damgaard, U.

Horm. Metab. Res. 20, 476-480, 1988

A;Title: Proinsulin heterogeneity in pigs.

A;Reference number: A60835; MUID:89032178; PMID:3181865

A;Accession: A60835

A;Molecule type: protein

A;Residues: 33-38,40-62 <SNE>

A;Note: the authors report the characterization of a connecting peptide variant lacking Ala-39

A;Accession: B60835

A;Molecule type: protein

A;Residues: 33-62 <SN2>

R;Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.

Adv. Protein Chem. 26, 279-402, 1972

A;Title: Insulin. the structure in the crystal and its reflection in chemistry and biology.

A;Reference number: A90017

A;Contents: annotation; X-ray crystallography, 1.9 angstroms

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-30/Domain: insulin chain B #status experimental <BCH>  
 F;1-30,64-84/Product: insulin #status experimental <MAT>  
 F;33-63/Domain: connecting peptide #status experimental <CPEP>  
 F;64-84/Domain: insulin chain A #status experimental <ACH>  
 F;7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match 82.7%; Score 383; DB 1; Length 84;  
 Best Local Similarity 86.0%; Pred. No. 1.1e-34;  
 Matches 74; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||||||||||||||||||||||||||||: || ||||| || || || |||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGG--GLGGLQALALEG 58

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||||||||||||||||||||||||||||
Db      59 PPQKRGIVEQCCTSICSLYQLENYCN 84
```

# RESULT 14

IPBO

insulin precursor - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 24-Apr-1984 #sequence\_revision 22-Apr-1995 #text\_change 16-Jul-1999

C;Accession: A40909; A92080; A92074; A91185; A90342; A90341; S48184; S48185;  
 S46258; A01585

R;D'Agostino, J.; Younes, M.A.; White, J.W.; Besch, P.K.; Field, J.B.; Frazier,  
 M.L.

Mol. Endocrinol. 1, 327-331, 1987

A;Title: Cloning and nucleotide sequence analysis of complementary  
 deoxyribonucleic acid for bovine preproinsulin.

A;Reference number: A40909; MUID:88288209; PMID:2456452

A;Accession: A40909

A;Molecule type: mRNA

A;Residues: 1-105 <DAA>

A;Cross-references: GB:M54979; NID:g163578; PIDN:AAA30722.1; PID:g163579

A;Experimental source: fetal pancreas

R;Nolan, C.; Margoliash, E.; Peterson, J.D.; Steiner, D.F.

J. Biol. Chem. 246, 2780-2795, 1971

A;Title: The structure of bovine proinsulin.

A;Reference number: A92080; MUID:71166442; PMID:4928892

A;Accession: A92080

A;Molecule type: protein

A;Residues: 25-105 <NOL>

R;Steiner, D.F.; Cho, S.; Oyer, P.E.; Terris, S.; Peterson, J.D.; Rubenstein,  
 A.H.

J. Biol. Chem. 246, 1365-1374, 1971

A;Title: Isolation and characterization of proinsulin C-peptide from bovine  
 pancreas.

A;Reference number: A92074; MUID:71116409; PMID:5545080

A;Accession: A92074

A;Molecule type: protein

A;Residues: 57-82 <STE>

R;Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.

Eur. J. Biochem. 20, 183-189, 1971

A;Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from  
 pancreas.

A;Reference number: A91185; MUID:71257721; PMID:5105368  
 A;Accession: A91185  
 A;Molecule type: protein  
 A;Residues: 57-82 <SAL>  
 R;Sanger, F.; Thompson, E.O.P.  
 Biochem. J. 53, 366-374, 1953  
 A;Title: The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates.  
 A;Reference number: A90342  
 A;Accession: A90342  
 A;Molecule type: protein  
 A;Residues: 85-105 <SAN>  
 R;Sanger, F.; Tuppy, H.  
 Biochem. J. 49, 481-490, 1951  
 A;Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates.  
 A;Reference number: A90341  
 A;Accession: A90341  
 A;Molecule type: protein  
 A;Residues: 25-54 <SA2>  
 R;Cheng, R.; Kawakishi, S.  
 Eur. J. Biochem. 223, 759-764, 1994  
 A;Title: Site-specific oxidation of histidine residues in glycated insulin mediated by Cu(2+).  
 A;Reference number: S48184; MUID:94333378; PMID:8055951  
 A;Accession: S48184  
 A;Molecule type: protein  
 A;Residues: 85-105 <CHE>  
 A;Accession: S48185  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 25-30,'X',32-42,'X',44-54 <CH2>  
 R;Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.  
 Biochem. J. 60, 541-556, 1955  
 A;Title: The disulphide bonds of insulin.  
 A;Reference number: A90343  
 A;Contents: annotation; amides; disulfides  
 R;Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.  
 FEBS Lett. 349, 205-209, 1994  
 A;Title: Existence of a molecular ruler in proteasomes suggested by analysis of degradation products.  
 A;Reference number: S46258; MUID:94326921; PMID:8050567  
 A;Accession: S46258  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 25-54 <WEN>  
 C;Superfamily: insulin  
 C;Keywords: hormone; pancreas  
 F;1-24/Domain: signal sequence #status predicted <SIG>  
 F;25-54/Domain: insulin chain B #status experimental <BCH>  
 F;25-54,85-105/Product: insulin #status experimental <MAT>  
 F;57-82/Domain: connecting peptide #status experimental <CPEP>  
 F;85-105/Domain: insulin chain A #status experimental <ACH>  
 F;31-91,43-104,90-95/Disulfide bonds: #status experimental

Query Match 79.2%; Score 366.5; DB 1; Length 105;  
 Best Local Similarity 80.2%; Pred. No. 8.9e-33;

```
Matches    69; Conservative      2; Mismatches     10; Indels       5; Gaps        1;
```

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|||||  
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPVGALELAGGPGAG-----GLEG 79

Qy 61 SLQKRGIQCCTSIQSLYQLENYCN 86  
|||  
Db 80 PPOKRGIVEOCCASVCSLYQLENYCN 105

F;88-108/Domain: insulin chain A #status experimental <ACH>  
F;31-94,43-107,93-98/Disulfide bonds: #status predicted

Query Match 79.0%; Score 366; DB 1; Length 108;  
Best Local Similarity 81.4%; Pred. No. 1e-32;  
Matches 70; Conservative 4; Mismatches 10; Indels 2; Gaps 1;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
      || ||||| |||||||||||||||||:|||| || || |:|||| | || || |||||
Db      25 FVKQHLCGPHLVEALYLVCGERGFFYTPKSRREVEDPQVEQLELGGSP--GDLQTLALEV 82

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
      : |||||:|||||||||||||||
Db      83 ARQKRGIVDQCCTSICSLYQLENYCN 108
```

Search completed: July 15, 2004, 16:37:33  
Job time : 10.7935 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:37:41 ; Search time 35.7799 Seconds  
(without alignments)  
751.267 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Match	Length	DB	ID	Description
No.						

1	463	100.0	86	9	US-09-878-380-1	Sequence 1, Appli
2	463	100.0	86	10	US-09-858-935B-4	Sequence 4, Appli
3	463	100.0	86	12	US-10-444-649-2	Sequence 2, Appli
4	463	100.0	86	12	US-10-444-701-2	Sequence 2, Appli
5	463	100.0	86	12	US-10-271-869-4	Sequence 4, Appli
6	463	100.0	86	13	US-10-028-410-2	Sequence 2, Appli
7	463	100.0	86	13	US-10-054-873-4	Sequence 4, Appli
8	463	100.0	86	14	US-10-444-326-2	Sequence 2, Appli
9	463	100.0	86	16	US-10-444-262-2	Sequence 2, Appli
10	463	100.0	96	9	US-09-947-563-4	Sequence 4, Appli
11	463	100.0	110	9	US-09-205-658-125	Sequence 125, App
12	463	100.0	110	9	US-09-815-229-3	Sequence 3, Appli
13	463	100.0	110	9	US-09-804-409A-9	Sequence 9, Appli
14	463	100.0	110	10	US-09-969-748C-6	Sequence 6, Appli
15	463	100.0	110	10	US-09-963-693-125	Sequence 125, App
16	463	100.0	110	12	US-10-411-037-44	Sequence 44, Appl
17	463	100.0	110	12	US-10-411-026-44	Sequence 44, Appl
18	463	100.0	110	14	US-10-038-686-1	Sequence 1, Appli
19	463	100.0	110	14	US-10-328-813-2	Sequence 2, Appli
20	463	100.0	110	15	US-10-383-285-2	Sequence 2, Appli
21	463	100.0	110	15	US-10-346-563-2	Sequence 2, Appli
22	463	100.0	110	15	US-10-321-717-2	Sequence 2, Appli
23	463	100.0	110	16	US-10-410-962-44	Sequence 44, Appl
24	463	100.0	110	16	US-10-411-049-44	Sequence 44, Appl
25	463	100.0	110	16	US-10-700-725-20	Sequence 20, Appl
26	463	100.0	110	16	US-10-410-930-44	Sequence 44, Appl
27	463	100.0	110	16	US-10-410-997-44	Sequence 44, Appl
28	463	100.0	110	16	US-10-411-012-44	Sequence 44, Appl
29	463	100.0	117	9	US-09-280-030-63	Sequence 63, Appl
30	463	100.0	130	9	US-09-280-030-62	Sequence 62, Appl
31	457	98.7	96	9	US-09-947-563-5	Sequence 5, Appli
32	438.5	94.7	124	15	US-10-221-677-24	Sequence 24, Appl
33	306	66.1	166	9	US-09-925-297-805	Sequence 805, App
34	300	64.8	56	9	US-09-815-229-10	Sequence 10, Appl
35	285	61.6	54	9	US-09-815-229-13	Sequence 13, Appl
36	267	57.7	52	13	US-10-054-873-5	Sequence 5, Appli
37	267	57.7	107	13	US-10-054-873-6	Sequence 6, Appli
38	267	57.7	137	16	US-10-101-454-39	Sequence 39, Appl
39	267	57.7	145	16	US-10-101-454-45	Sequence 45, Appl
40	267	57.7	146	16	US-10-101-454-48	Sequence 48, Appl
41	267	57.7	150	13	US-10-054-873-7	Sequence 7, Appli
42	263.5	56.9	102	16	US-10-101-454-36	Sequence 36, Appl
43	261.5	56.5	51	10	US-09-858-935B-5	Sequence 5, Appli
44	261.5	56.5	51	12	US-10-444-649-3	Sequence 3, Appli
45	261.5	56.5	51	12	US-10-444-701-3	Sequence 3, Appli

#### ALIGNMENTS

RESULT 1  
 US-09-878-380-1  
 ; Sequence 1, Application US/09878380  
 ; Patent No. US20020160435A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fujirebio Inc.



```
; APPLICANT: KITAJIMA, Sachiko
; APPLICANT: KURANO, Yoshihiro
; APPLICANT: NAKATSUBO, Kaoru
; APPLICANT: NISHIZONO, Isao
; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit
Therefor
; FILE REFERENCE: 0760-0291P
; CURRENT APPLICATION NUMBER: US/09/878,380
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: JP 2000-174691
; PRIOR FILING DATE: 2000-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-878-380-1
```

```
Query Match          100.0%; Score 463; DB 9; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.1e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
```

## RESULT 2

```
US-09-858-935B-4
; Sequence 4, Application US/09858935B
; Publication No. US20030069177A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 4
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-935B-4
```

```
Query Match          100.0%; Score 463; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.1e-44;
```

```

Matches      86;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy           1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy           61 SLQKRGIVEQCCTSICSLYQLENYCN 86
              ||||||||||||||||||||||||
Db           61 SLQKRGIVEQCCTSICSLYQLENYCN 86

```

RESULT 3

```

US-10-444-649-2
; Sequence 2, Application US/10444649
; Publication No. US20040033951A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,649
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/724,479
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-649-2

```

```

Query Match          100.0%;  Score 463;  DB 12;  Length 86;
Best Local Similarity 100.0%;  Pred. No. 2.1e-44;
Matches      86;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy           1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy           61 SLQKRGIVEQCCTSICSLYQLENYCN 86
              ||||||||||||||||||||||||
Db           61 SLQKRGIVEQCCTSICSLYQLENYCN 86

```

RESULT 4

```

US-10-444-701-2
; Sequence 2, Application US/10444701
; Publication No. US20040033952A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,701
; CURRENT FILING DATE: 2003-05-22

```

; PRIOR APPLICATION NUMBER: US/09/723,866  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: US/09/477,923  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-444-701-2

Query Match 100.0%; Score 463; DB 12; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.1e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 5

US-10-271-869-4

; Sequence 4, Application US/10271869  
; Publication No. US20030211992A1  
; GENERAL INFORMATION:  
; APPLICANT: Dubaquie, Yves  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Lowman, Henry B.  
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
; FILE REFERENCE: P1794R1  
; CURRENT APPLICATION NUMBER: US/10/271,869  
; CURRENT FILING DATE: 2002-10-16  
; PRIOR APPLICATION NUMBER: US/09/858,935  
; PRIOR FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: US 60/248,985  
; PRIOR FILING DATE: 2000-11-15  
; PRIOR APPLICATION NUMBER: US 60/204,490  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 153  
; SEQ ID NO 4  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-271-869-4

Query Match 100.0%; Score 463; DB 12; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.1e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 6

US-10-028-410-2  
 ; Sequence 2, Application US/10028410  
 ; Publication No. US20020160955A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dubaquié, Yves  
 ; APPLICANT: Lowman, Henry  
 ; TITLE OF INVENTION: PROTEIN VARIANTS  
 ; FILE REFERENCE: P1712R1-1  
 ; CURRENT APPLICATION NUMBER: US/10/028,410  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: US/09/477,924  
 ; PRIOR FILING DATE: 2000-01-05  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SEQ ID NO 2  
 ; LENGTH: 86  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-028-410-2

Query Match 100.0%; Score 463; DB 13; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-44;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |||  
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |||  
 Db 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

RESULT 7

US-10-054-873-4  
 ; Sequence 4, Application US/10054873  
 ; Publication No. US20020164712A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gan, Zhong Ru  
 ; TITLE OF INVENTION: Chimeric Protein Containing an  
 ; Intramolecular Chaperone-Like Sequence  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend and Crew LLP  
 ; STREET: Two Embarcadero Center, Eighth Floor  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible

```

;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/10/054,873
;      FILING DATE: 22-Jan-2002
;      CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: WO PCT/CN98/00052
;      FILING DATE: 31-MAR-1998
;      APPLICATION NUMBER: US 09/423,100
;      FILING DATE: 11-DEC-2000
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Mycroft, Frank J
;      REGISTRATION NUMBER: 46,946
;      REFERENCE/DOCKET NUMBER: 020167-000130US
;      INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 86 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: <Unknown>
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-054-873-4

```

```

Query Match          100.0%; Score 463; DB 13; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.1e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
        ||||||||||||||||||||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86

```

# RESULT 8

US-10-444-326-2

```

; Sequence 2, Application US/10444326
; Publication No. US20030191065A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquier, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,326
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/723,866
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT

```

; ORGANISM: Homo sapiens  
US-10-444-326-2

Query Match 100.0%; Score 463; DB 14; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.1e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          ||||||||||||||||||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
```

RESULT 9

US-10-444-262-2  
; Sequence 2, Application US/10444262  
; Publication No. US20040023883A1  
; GENERAL INFORMATION:  
; APPLICANT: Dubaquier, Yves  
; APPLICANT: Lowman, Henry  
; TITLE OF INVENTION: PROTEIN VARIANTS  
; FILE REFERENCE: P1712R1  
; CURRENT APPLICATION NUMBER: US/10/444,262  
; CURRENT FILING DATE: 2003-05-22  
; PRIOR APPLICATION NUMBER: US/09/724,478  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: US/09/477,923  
; PRIOR FILING DATE: 2000-01-05  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-444-262-2

Query Match 100.0%; Score 463; DB 16; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.1e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          ||||||||||||||||||
Db      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
```

RESULT 10

US-09-947-563-4  
; Sequence 4, Application US/09947563  
; Patent No. US20020156234A1  
; GENERAL INFORMATION:  
; APPLICANT: Rubroder, Franz-Josef

```

;           Keller, Reinhold
;   TITLE OF INVENTION: Improved process for obtaining
;                       insulin precursors having correctly bonded cystine
bridges
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
;               Dunner
;       STREET: 1300 I Street, N.W.
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/947,563
;       FILING DATE: 07-Sep-2001
;       CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 09/134,836
;       FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Leslie McDonell
;       REGISTRATION NUMBER: 34,872
;       REFERENCE/DOCKET NUMBER: 02481.1600-00000
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (202) 408-4000
;       TELEFAX: (202) 408-4400
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 96 amino acids
;           TYPE: amino acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       ORIGINAL SOURCE:
;           ORGANISM: Escherichia coli
;       FEATURE:
;           NAME/KEY: Protein
;           LOCATION: 1..96
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4

```

```

Query Match          100.0%; Score 463; DB 9; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.4e-44;
Matches   86; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 70
          ||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy     61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          ||||||||||||||||||||||||

```

175

US-09-205-658-125

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.8e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 12

US-09-815-229-3

```
; Sequence 3, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS
DISORDERS
; FILE REFERENCE: P1786R1US
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
```



; SEQ ID NO 3  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-815-229-3

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.8e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 13

US-09-804-409A-9

; Sequence 9, Application US/09804409A  
; Patent No. US20020155100A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEFFER, TIMOTHY J.  
; APPLICANT: CHEUNG, ANTHONY T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN  
; TITLE OF INVENTION: EXPRESSION IN GUT  
; FILE REFERENCE: 029996/027 8721  
; CURRENT APPLICATION NUMBER: US/09/804,409A  
; CURRENT FILING DATE: 2001-03-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-804-409A-9

Query Match 100.0%; Score 463; DB 9; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.8e-44;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|  
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
  
Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
|  
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 14

US-09-969-748C-6

; Sequence 6, Application US/09969748C  
; Publication No. US20030161809A1  
; GENERAL INFORMATION:

```
; APPLICANT: ARIZEKE PHARMACEUTICALS, INC.
; APPLICANT: HOUSTON, Lou, L.
; APPLICANT: SHERIDAN, Philip, J.
; APPLICANT: HAWLEY, Stephen
; APPLICANT: GLYNN, Jacqueline, M.
; APPLICANT: CHAPIN, Steven
; APPLICANT: BASU, Amaresh
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TRANSPORT OF
BIOLOGICALLY ACTIVE
; TITLE OF INVENTION: AGENTS ACROSS CELLULAR BARRIERS
; FILE REFERENCE: 057220-0303
; CURRENT APPLICATION NUMBER: US/09/969,748C
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/267,601
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/248,819
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: US 60/248,478
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/237,929
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-748C-6
```

```
Query Match          100.0%; Score 463; DB 10; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.8e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

# RESULT 15

US-09-963-693-125

```
; Sequence 125, Application US/09963693
; Publication No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
```

```
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 125
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-693-125
```

```
Query Match          100.0%; Score 463; DB 10; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.8e-44;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
        ||||||||||||||||||||||||||||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

```
Search completed: July 15, 2004, 17:05:08
Job time : 36.7799 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:50 ; Search time 29.3619 Seconds  
(without alignments)  
924.141 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaphage:\*  
17: sp\_archaeophages:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

---

1	463	100.0	110	6	Q8HXV2	Q8hvx2 pongo pygma
2	388	83.8	110	6	Q8WNW6	Q8wnw6 felis silve
3	342	73.9	65	6	Q8HZ81	Q8hz81 gorilla gor
4	342	73.9	65	6	Q8HZ80	Q8hz80 pongo pygma
5	246.5	53.2	106	13	Q9I8Q7	Q9i8q7 rana pipien
6	235.5	50.9	111	13	Q98TA7	Q98ta7 osteoglossu
7	230.5	49.8	110	13	Q98TA8	Q98ta8 pantodon bu
8	222.5	48.1	110	13	Q90ZY1	Q90zyl hiodon alos
9	219	47.3	111	13	Q98TB0	Q98tb0 chitala chi
10	214.5	46.3	108	13	Q9DDE5	Q9dde5 brachydanio
11	212.5	45.9	108	13	Q90ZN4	Q90zn4 catla catla
12	210.5	45.5	87	13	Q98TA9	Q98ta9 gnathonemus
13	205.5	44.4	108	13	Q98TB1	Q98tb1 catostomus
14	203.5	44.0	91	13	Q98TB2	Q98tb2 ambloplites
15	189	40.8	41	11	Q62543	Q62543 mus spretus
16	162	35.0	39	11	Q62542	Q62542 mus spretus
17	142.5	30.8	104	13	Q7T107	Q7t107 dicentrarch
18	142.5	30.8	108	13	Q800N0	Q800n0 morone chry
19	142.5	30.8	108	13	Q800M9	Q800m9 morone saxa
20	142.5	30.8	108	13	Q800M8	Q800m8 morone chry
21	142.5	30.8	108	13	Q800M7	Q800m7 morone amer
22	142.5	30.8	159	13	O93607	O93607 paralichthy
23	142.5	30.8	182	13	O73720	O73720 oreochromis
24	142.5	30.8	182	13	O42289	O42289 oreochromis
25	142.5	30.8	182	13	P79824	P79824 oreochromis
26	142.5	30.8	185	13	O57436	O57436 paralichthy
27	142.5	30.8	186	13	O93527	O93527 paralichthy
28	142.5	30.8	186	13	Q7T1A7	Q7t1a7 perca flave
29	141.5	30.6	186	13	Q800Y5	Q800y5 siganus gut
30	141	30.5	207	13	Q90XD0	Q90xd0 cyprinus ca
31	140.5	30.3	132	13	Q8AV14	Q8av14 petromyzon
32	138.5	29.9	153	13	O93380	O93380 meleagris g
33	137	29.6	185	13	Q9YI57	Q9yi57 acanthopagr
34	137	29.6	210	13	Q91443	Q91443 squalus aca
35	136.5	29.5	62	13	Q9IAA0	Q9iaa0 carassius a
36	136.5	29.5	116	13	Q91161	Q91161 oncorhynchu
37	136.5	29.5	117	13	Q91476	Q91476 salmo salar
38	136.5	29.5	145	13	Q91475	Q91475 salmo salar
39	136.5	29.5	149	13	Q91231	Q91231 oncorhynchu
40	136.5	29.5	155	13	Q91162	Q91162 oncorhynchu
41	136.5	29.5	161	13	Q91230	Q91230 oncorhynchu
42	136.5	29.5	188	13	P81268	P81268 oncorhynchu
43	136.5	29.5	188	13	Q91965	Q91965 oncorhynchu
44	136	29.4	215	13	Q800Y4	Q800y4 siganus gut
45	135.5	29.3	184	13	O42336	O42336 myoxocephal

# ALIGNMENTS

## RESULT 1

Q8HXV2

ID Q8HXV2 PRELIMINARY; PRT; 110 AA.  
AC Q8HXV2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Insulin precursor.  
 GN INS.  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
 OX NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stead J.D.H., Jeffreys A.J.;  
 RT "Haplotype diversity at the insulin region."  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY137503; AAN06937.1; -.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA; 12038 MW; 22D2B32B94F520F8 CRC64;

Query Match 100.0%; Score 463; DB 6; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-46;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |||  
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

## RESULT 2

### Q8WNW6

ID Q8WNW6 PRELIMINARY; PRT; 110 AA.  
 AC Q8WNW6;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Okamoto S., Morimatsu M.;  
 RT "cat insulin."  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AB043535; BAB84110.1; -.  
 DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 SQ SEQUENCE 110 AA; 12069 MW; 95FB6E170C7BECA4 CRC64;

Query Match 83.8%; Score 388; DB 6; Length 110;  
 Best Local Similarity 83.7%; Pred. No. 1.4e-37;  
 Matches 72; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||||||||||||||||| ||| ||| ||| |||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAEDLQKDAELGEAPGAGGLQPSALEA 84  
 Qy 61 SLQKRGIVEQCCTSIQSLYQLENYCN 86  
 ||||||||| |:|||||:|  
 Db 85 PLQKRGIVEQCCASVCSLYQLEHYCN 110

# RESULT 3

Q8HZ81

ID Q8HZ81 PRELIMINARY; PRT; 65 AA.  
 AC Q8HZ81;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Insulin (Fragment).  
 OS Gorilla gorilla (gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 OX NCBI\_TaxID=9593;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA O'hUigin C., Tichy H., Klein J.;  
 RT "Molecular evolution in higher primates; gene specific and organism  
 RT specific characteristics."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY092023; AAM76640.1; -.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR SMART; SM00078; IIGF; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 65 65  
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 73.9%; Score 342; DB 6; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRG 66  
 |||||||||||||||||||||||||||||||||||

Db 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRG 60

Qy 67 IVEQC 71  
 |||||

Db 61 IVEQC 65

# RESULT 4

Q8HZ80

ID Q8HZ80 PRELIMINARY; PRT; 65 AA.  
 AC Q8HZ80;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Insulin (Fragment).  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
 OX NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA O'hUigin C., Tichy H., Klein J.;  
 RT "Molecular evolution in higher primates; gene specific and organism  
 RT specific characteristics."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY092024; AAM76641.1; -.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR SMART; SM00078; IIGF; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 65 65  
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 73.9%; Score 342; DB 6; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-32;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRG 66  
 |||||

Db 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRG 60

Qy 67 IVEQC 71  
 |||||

Db 61 IVEQC 65

# RESULT 5

Q9I8Q7

ID Q9I8Q7 PRELIMINARY; PRT; 106 AA.  
 AC Q9I8Q7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin.



OS Rana pipiens (Northern leopard frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.  
 OX NCBI\_TaxID=8404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20362507; PubMed=10818274;  
 RA Irwin D.M., Sivarajah P.;  
 RT "Proinsulin cDNAs from the leopard frog, Rana pipiens: evolution of  
 RT proinsulin processing."  
 RL Comp. Biochem. Physiol. 125B:405-410(2000).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF227187; AAF87285.1; -.  
 DR HSSP; P01315; 1SDB.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 SQ SEQUENCE 106 AA; 12183 MW; 3A870EEC70217F92 CRC64;

Query Match 53.2%; Score 246.5; DB 13; Length 106;  
 Best Local Similarity 51.5%; Pred. No. 4.9e-21;  
 Matches 52; Conservative 9; Mismatches 7; Indels 33; Gaps 4;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPL--AL 58  
 | ||:|||||:||||:||||:|:|:| | ||| |  
 Db 24 FDNQYLCGSHLVEALYMVCGRGFFYSPRSRRDLE-----QPLVNGL 65  
 QY 59 EGS-----LQKR--GIVEQCCTSICSLYQLENYCN 86  
 :|| ||| ||||| : |||| |||||  
 Db 66 QGSELDEMQVQSQAQKRPKPGIVEQCCHNTCSLYDLENYCN 106

# RESULT 6

Q98TA7

ID Q98TA7 PRELIMINARY; PRT; 111 AA.  
 AC Q98TA7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Osteoglossum bicirrhosum (silver arawana).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Osteoglossidae; Osteoglossum.  
 OX NCBI\_TaxID=109271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid.";

RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199589; AAK28713.1; -.  
 DR HSSP; P01315; 1MPJ.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; ILGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 12491 MW; AC9E19D2D4866D20 CRC64;

Query Match 50.9%; Score 235.5; DB 13; Length 111;  
 Best Local Similarity 54.1%; Pred. No. 9.9e-20;  
 Matches 46; Conservative 12; Mismatches 26; Indels 1; Gaps 1;

Qy 3 NQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSL 62  
 :| |||||:||||:||||:||||:||||:| | | | | :| |  
 Db 27 SQRLCGSHLVDALYMVCGDRGFFYSPKSRREAPELLGLSPKSGQENEVDEYPYKEQGEL 86  
  
 Qy 63 Q-KRGIVEQCCTSICSLYQLENYCN 86  
 : ||||| || |:: |:|||  
 Db 87 KVKRGIVEQCCHRPCNIFDLQNYCN 111

#### RESULT 7

Q98TA8

ID Q98TA8 PRELIMINARY; PRT; 110 AA.  
 AC Q98TA8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin.  
 OS Pantodon buchholtzi (Butterflyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Pantodontidae; Pantodon.  
 OX NCBI\_TaxID=8276;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199588; AAK28712.1; -.  
 DR HSSP; P01308; 1HIS.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.

DR Pfam; PF00049; Insulin; 1.  
DR PRINTS; PR00277; INSULINB.  
DR SMART; SM00078; IIGF; 1.  
DR PROSITE; PS00262; INSULIN; 1.  
SO SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

```
Qy      3 NQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLAL---- 58  
       :|||||||:|||:|||:|||:|||::| | | | :  
Db     26 SQHLCGSHLVLDALYMVCGEKGFYQPKTKRDVD-----PLLGLSPKSAQENE 73  
  
Qy     59 -----EGSLQ-KRGIVEQCCTSICSLYQLENYCN 86  
       :| |: ||||| | | | :: |: ||| |  
Db    74 ADEYPYKDQGDLKVKRGIVEQCCHPCNIFDLQNYCN 110
```

## RESULT 8

090ZY1

```

ID      Q90ZY1      PRELIMINARY;      PRT;      110 AA.
AC      Q90ZY1;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Preproinsulin (Fragment).
OS      Hiodon alosoides (goldeye).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC      Osteoglossiformes; Hiodontidae; Hiodon.
OX      NCBI_TaxID=54904;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21203577; PubMed=11306171;
RA      Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT      "Molecular cloning of preproinsulin cDNAs from several
RT      osteoglossomorphs and a cyprinid.";
RL      Mol. Cell. Endocrinol. 174:51-58(2001).
CC      -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR      EMBL; AF282408; AAK54684.1; -.
DR      HSSP; P01308; 1LNP.
DR      GO; GO:0005576; C:extracellular; IEA.
DR      GO; GO:0005179; F:hormone activity; IEA.
DR      GO; GO:0007582; P:physiological processes; IEA.
DR      InterPro; IPR004825; Ins/IGF/relax.
DR      Pfam; PF00049; Insulin; 1.
DR      PRINTS; PR00277; INSULINB.
DR      SMART; SM00078; IIGF; 1.
DR      PROSITE; PS00262; INSULIN; 1.
FT      NON_TER      110      110
SQ      SEQUENCE      110 AA;      12343 MW;      BDECCD7703E52E06 CRC64;

```

Query Match 48.1%; Score 222.5; DB 13; Length 110;  
Best Local Similarity 45.4%; Pred. No. 3.3e-18;  
Matches 44; Conservative 13; Mismatches 15; Indels 25; Gaps 3;

Qy 3 NQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLAL---- 58  
 :|||||||:||||:||||:||||:| | | :  
 Db 26 SQHLCGSHLVDALYMVCGEKGGFFYQPKTKRDVD-----PLLGFLLSPKSAQENE 73

Qy 59 -----EGSLQ-KRGIVEQCCTSICSLYQLENYCN 86  
 :| |: ||||| |:: | ||  
 Db 74 ADEYPYKDGDLKVKRGIVEQCCHRPCNIFDLNQYCN 110

# RESULT 9

Q98TB0

ID Q98TB0 PRELIMINARY; PRT; 111 AA.  
 AC Q98TB0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Chitala chitala (clown knifefish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Notopteridae; Chitala.  
 OX NCBI\_TaxID=112163;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199586; AAK28710.1; -.  
 DR HSSP; P01308; 1LPH.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 47.3%; Score 219; DB 13; Length 111;  
 Best Local Similarity 49.0%; Pred. No. 8.5e-18;  
 Matches 48; Conservative 7; Mismatches 17; Indels 26; Gaps 4;

Qy 3 NQHLCGSHLVEALYLVCGERGFFYTPK-TRREAEDLQVGQVELGGGPGAGSLQPLA-LEG 60  
 |||||||:||||:||||:||||:| | | : ||  
 Db 26 NQHLCGSHLVEALYLVCGERGFFYNPKMDKRD AE-----PLLGFLLSPKSGLEN 73

Qy 61 SL-----QKRGIVEQCCTSICSLYQLENYCN 86  
 : ||||| |:: | ||  
 Db 74 EVDEYPFKDQGDVKMKRGIVEQCCHRPCNIFDQNYCN 111

09DDE5

Query Match 46.3%; Score 214.5; DB 13; Length 108;  
Best Local Similarity 45.8%; Pred. No. 2.8e-17;  
Matches 44; Conservative 11; Mismatches 14; Indels 27; Gaps 3;

```
QY      4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGPGAGSLQPLALEGS-- 61  
        |||||::||| | | | : | | | : ::  
Db     27 QHLCGSHLVLDALYLVCGP TGFFYNPK--RDVE-----PLLGLFPKSAQETEV 72  
  
QY     62 -----LQKR GIVEQCCTSI CSLYQL ENYCN 86  
        ::||| | | | | | : : : | | |  
Db    73 ADFAFKDHAELIR KRGIVEQCCHKPCSIFELQNYCN 108
```

## Q90ZN4

Q90ZN4  
ID Q90ZN4 PRELIMINARY; PRT; 108 AA.  
AC Q90ZN4;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin.  
 OS *Catla catla* (catla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; *Catla*.  
 OX NCBI\_TaxID=72446;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,  
 RA Bandyopadhyaya I., Wakabayashi K.;  
 RT "A new cell secreting insulin."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF373021; AAK51558.1; -.  
 DR HSSP; P01308; 1LNP.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5D59 CRC64;

Query Match 45.9%; Score 212.5; DB 13; Length 108;  
 Best Local Similarity 44.8%; Pred. No. 4.7e-17;  
 Matches 43; Conservative 12; Mismatches 14; Indels 27; Gaps 3;

QY 4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGS-- 61  
 |||||:||||| ||| | |: : | | | : : :  
 Db 27 QHLCGSHLVDALYLVCGPTGFFYNPK--RDVDPLM-----GFLPPKSAQETEV 72  
 QY 62 -----LQKRGIVEQCCTSICSLYQLENYCN 86  
 :|||:||||| ||:::|:|  
 Db 73 ADFAFKDHAEVIRKRGIVEQCCHKPCSIFELQNYCN 108

# RESULT 12

Q98TA9

ID Q98TA9 PRELIMINARY; PRT; 87 AA.

AC Q98TA9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Preproinsulin (Fragment).

OS *Gnathonemus petersii*.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;

OC Osteoglossiformes; Mormyridae; *Gnathonemus*.

OX NCBI\_TaxID=42645;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199587; AAK28711.1; -.  
 DR HSSP; P01308; 1HIS.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 87 87  
 SQ SEQUENCE 87 AA; 9874 MW; FF448ED35D2453F5 CRC64;

Query Match 45.5%; Score 210.5; DB 13; Length 87;  
 Best Local Similarity 50.6%; Pred. No. 6.3e-17;  
 Matches 43; Conservative 11; Mismatches 28; Indels 3; Gaps 2;

QY 4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGP--GAGSLQPLALEGS 61  
 |||||:|||||:| |:|: |:|: || | :  
 Db 4 QHLCGSHLVEALFLVCGERGFFNPDTRDVSLLGFLSPKSGPENEADEYRYKEQAEV 62  
 QY 62 LQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||| |::| |||  
 Db 63 KVKRGIVEQCCHHPCNIFDLNQYCN 87

# RESULT 13

Q98TB1

ID Q98TB1 PRELIMINARY; PRT; 108 AA.  
 AC Q98TB1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Catostomus commersoni (White sucker).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Catostomidae; Catostomus.  
 OX NCBI\_TaxID=7971;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21203577; PubMed=11306171;  
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNAs from several  
 RT osteoglossomorphs and a cyprinid."  
 RL Mol. Cell. Endocrinol. 174:51-58(2001).  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199585; AAK28709.1; -.

DR HSSP; P01308; 1LPH.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11873 MW; E426310696FBAFC8 CRC64;

Query Match 44.4%; Score 205.5; DB 13; Length 108;  
 Best Local Similarity 50.0%; Pred. No. 3.1e-16;  
 Matches 43; Conservative 12; Mismatches 24; Indels 7; Gaps 4;

Qy 4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGS-- 61  
 |||||:||||| ||| || |: : |: : || : | :  
 Db 27 QHLCGSHLVDALYLVCGPTGFFYNPK--RDVDPL-IGFLPPKSGP-ENEVADFAFKDHAE 82  
 Qy 62 -LQKRGIVEQCCTSICSLYQLENYCN 86  
 ::||||||| |:: || |||  
 Db 83 LIRKRGIVEQCCHRPCNIFDLEKYCN 108

#### RESULT 14

Q98TB2

ID Q98TB2 PRELIMINARY; PRT; 91 AA.  
 AC Q98TB2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Preproinsulin (Fragment).  
 OS Ambloplites rupestris (Rock bass).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;  
 OC Centrarchidae; Ambloplites.  
 OX NCBI\_TaxID=109273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Al-Mahrouki A.A., Irwin D.M., Youson J.H.;  
 RT "Molecular cloning of preproinsulin cDNA from the rock bass."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
 DR EMBL; AF199584; AAK28708.1; -.  
 DR HSSP; P01308; 1LPH.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007582; P:physiological processes; IEA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 FT NON\_TER 1 1



FT NON\_TER 91 91  
SQ SEQUENCE 91 AA; 10100 MW; E86C8B256DC69D39 CRC64;

Query Match 44.0%; Score 203.5; DB 13; Length 91;  
Best Local Similarity 46.7%; Pred. No. 4.4e-16;  
Matches 42; Conservative 13; Mismatches 26; Indels 9; Gaps 4;

Qy 4 QHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQ---VGQVELGGGPGA-GSLQPLALE 59  
| | | | | : | | | | | : | | | | : | : | : | : | : | :  
Db 4 QHLCGSHLVDALYLVCGRGFFYNPK--RDVDPLMGFLPPKADGAAAPGGENEVAEFAFK 61  
  
Qy 60 GSLQ---KRGIVEQCCTSICSLYQLENYCN 86  
: : | | | | | | | | | : : : | | |  
Db 62 DQMEMMVKRGIVEQCCHPCNIFDLGRYCN 91

RESULT 15

Q62543

ID Q62543 PRELIMINARY; PRT; 41 AA.  
AC Q62543;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Insulin 2 (Fragment).  
GN INS2.  
OS Mus spretus (Western wild mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRET/EI;  
RX MEDLINE=94319082; PubMed=8043949;  
RA Ko M.S., Wang X., Horton J.H., Hagen M.D., Takahashi N., Maezaki Y.,  
RA Nadeau J.H.;  
RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";  
RL Mamm. Genome 5:349-355(1994).  
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR EMBL; U05730; AAB60474.1; -.  
DR PIR; I49419; I49419.  
DR HSSP; P01308; 1A7F.  
DR MGD; MGI:96573; Ins2.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0005179; F:hormone activity; IEA.  
DR GO; GO:0006006; P:glucose metabolism; IEA.  
DR GO; GO:0007582; P:physiological processes; IEA.  
DR InterPro; IPR004825; Ins/IGF/relax.  
DR Pfam; PF00049; Insulin; 1.  
DR PRINTS; PR00276; INSULINA.  
DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.  
KW Hormone; Glucose metabolism; Multigene family.  
FT NON\_TER 1 1  
FT CHAIN 21 41 A CHAIN.  
SQ SEQUENCE 41 AA; 4361 MW; 55CDB871FF720672 CRC64;

Query Match 40.8%; Score 189; DB 11; Length 41;  
Best Local Similarity 85.4%; Pred. No. 8.9e-15;  
Matches 35; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 46 GGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN 86  
||||| || |||| : |||||:|||||||  
Db 1 GGPGAGDLQTLALEVAQQKRGIVDQCCTSICSLYQLENYCN 41

Search completed: July 15, 2004, 16:40:56  
Job time : 37.5286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:28:49 ; Search time 5.93657 Seconds  
(without alignments)  
754.314 Million cell updates/sec

Title: US-09-423-100-4  
Perfect score: 463  
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSIICSLYQLENYCN 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	463	100.0	110	1	INS_HUMAN	P01308	homo sapien
2	463	100.0	110	1	INS_PANTR	P30410	pan troglod
3	456	98.5	110	1	INS_CERAE	P30407	cercopithec
4	456	98.5	110	1	INS_MACFA	P30406	macaca fasc
5	424	91.6	110	1	INS_RABIT	P01311	oryctolagus
6	417	90.1	110	1	INS_CANFA	P01321	canis famil
7	413	89.2	110	1	INS_SPETR	Q91xi3	spermophilu
8	394	85.1	86	1	INS_HORSE	P01310	equus cabal
9	394	85.1	110	1	INS2_MOUSE	P01326	mus musculu
10	394	85.1	110	1	INS2_RAT	P01323	rattus norv
11	392	84.7	108	1	INS_AOTTR	P10604	aotus trivi
12	392	84.7	110	1	INS_CRILO	P01313	cricetulus
13	385	83.2	110	1	INS1_RAT	P01322	rattus norv
14	383	82.7	108	1	INS_PIG	P01315	sus scrofa
15	377	81.4	110	1	INS_PSAOB	Q62587	psammomys o
16	366.5	79.2	105	1	INS_BOVIN	P01317	bos taurus
17	366	79.0	108	1	INS1_MOUSE	P01325	mus musculu

18	362.5	78.3	105	1	INS_SHEEP	P01318	ovis aries
19	334.5	72.2	108	1	INS_RODSP	P21563	rodentia sp
20	314	67.8	110	1	INS_CAVPO	P01329	cavia porce
21	277.5	59.9	109	1	INS_OCTDE	P17715	octodon deg
22	276.5	59.7	103	1	INS_SELRF	P51463	selasphorus
23	265.5	57.3	106	1	INS2_XENLA	P12707	xenopus lae
24	265.5	57.3	107	1	INS_CHICK	P01332	gallus gall
25	262.5	56.7	106	1	INS1_XENLA	P12706	xenopus lae
26	256.5	55.4	51	1	INS_BALPH	P01312	balaenopter
27	256.5	55.4	51	1	INS_ELEMA	P01316	elephas max
28	256.5	55.4	81	1	INS_ANAPL	P01333	anas platyr
29	251.5	54.3	51	1	INS_ACOCA	P01324	acomys cahi
30	246.5	53.2	51	1	INS_BALBO	P01314	balaenopter
31	246.5	53.2	51	1	INS_CAMDR	P01320	camelus dro
32	246.5	53.2	51	1	INS_CAPHI	P01319	capra hircu
33	245.5	53.0	51	1	INS_FELCA	P06306	felis silve
34	239.5	51.7	51	1	INS_DIDMA	P18109	didelphis m
35	234.5	50.6	51	1	INS_CHIBR	P01327	chinchilla
36	231.5	50.0	51	1	INS_ANSAN	P07454	anser anser
37	228	49.2	115	1	INS_VERMO	Q9w7r2	verasper mo
38	227.5	49.1	51	1	INS_HYSCR	P01328	hystrix cri
39	227.5	49.1	51	1	INS_TRASC	P31887	trachemys s
40	225	48.6	52	1	INS_ACIGU	P81423	acipenser g
41	221.5	47.8	116	1	INS_LOPPI	P01341	lophius pis
42	218.5	47.2	51	1	INS_ORNAN	Q9tqy7	ornithorhyn
43	218.5	47.2	105	1	INS_ONCKE	P04667	oncorhynchu
44	216.5	46.8	89	1	INS_CALMI	P13190	callorhynch
45	215	46.4	113	1	INS_ORENI	P81025	oreochromis

# ALIGNMENTS

## RESULT 1

### INS\_HUMAN

ID INS\_HUMAN STANDARD; PRT; 110 AA.  
AC P01308;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=80120725; PubMed=6243748;  
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,  
RA Goodman H.M.;  
RT "Sequence of the human insulin gene."  
RL Nature 284:26-32(1980).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=80236313; PubMed=6248962;  
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;

RT "Genetic variation in the human insulin gene."  
 RL Science 209:612-615(1980).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=80054779; PubMed=503234;  
 RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,  
 RA Rutter W.J.;  
 RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin."  
 RL Nature 282:525-527(1979).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=80147417; PubMed=6927840;  
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;  
 RT "Nucleotide sequence of human preproinsulin complementary DNA."  
 RL Science 208:57-59(1980).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93364428; PubMed=8358440;  
 RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;  
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1  
 RT kb segment of DNA spanning the insulin gene and associated VNTR."  
 RL Nat. Genet. 4:305-310(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 1-59 FROM N.A.  
 RC TISSUE=Blood;  
 RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;  
 RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)  
 RT within the 5' region of insulin gene."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RA Nicol D.S.H.W., Smith L.F.;  
 RT "Amino-acid sequence of human insulin."

RL Nature 187:483-485(1960).  
 RN [9]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE=71116410; PubMed=5101771;  
 RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;  
 RT "Studies on human proinsulin. Isolation and amino acid sequence of  
 RT the human pancreatic C-peptide.";  
 RL J. Biol. Chem. 246:1375-1386(1971).  
 RN [10]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE=71257722; PubMed=5560404;  
 RA Ko A., Smyth D.G., Markussen J., Sundby F.;  
 RT "The amino acid sequence of the C-peptide of human proinsulin.";  
 RL Eur. J. Biochem. 20:190-199(1971).  
 RN [11]  
 RP SYNTHESIS.  
 RX MEDLINE=75077277; PubMed=4443293;  
 RA Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;  
 RT "Total synthesis of human insulin under directed formation of the  
 RT disulfide bonds.";  
 RL Helv. Chim. Acta 57:2617-2621(1974).  
 RN [12]  
 RP SYNTHESIS OF 57-87.  
 RX MEDLINE=75040007; PubMed=4803504;  
 RA Naithani V.K.;  
 RT "Studies on polypeptides, IV. The synthesis of C-peptide of human  
 RT proinsulin.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).  
 RN [13]  
 RP SYNTHESIS OF 65-69 AND 70-73.  
 RX MEDLINE=73161263; PubMed=4698555;  
 RA Geiger R., Volk A.;  
 RT "Synthesis of peptides with the properties of human proinsulin C  
 RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13  
 RT of human proinsulin C peptides.";  
 RL Chem. Ber. 106:199-205(1973).  
 RN [14]  
 RP SYNTHESIS OF 84-87.  
 RX MEDLINE=73161261; PubMed=4698553;  
 RA Geiger R., Jaeger G., Keonig W., Treuth G.;  
 RT "Synthesis of peptides with the properties of human proinsulin C  
 RT peptides (hC peptide). I. Scheme for the synthesis and preparation of  
 RT the sequence 28-31 of human proinsulin C peptide.";  
 RL Chem. Ber. 106:188-192(1973).  
 RN [15]  
 RP VARIANT LOS ANGELES SER-48.  
 RX MEDLINE=84016053; PubMed=6312455;  
 RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;  
 RT "Studies on mutant human insulin genes: identification and sequence  
 RT analysis of a gene encoding [SerB24]insulin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).  
 RN [16]  
 RP VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.  
 RX MEDLINE=84170233; PubMed=6424111;  
 RA Shoelson S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,  
 RA Rubenstein A.H., Tager H.;  
 RT "Identification of a mutant human insulin predicted to contain a

RT serine-for-phenylalanine substitution.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).  
 RN [17]  
 RP VARIANT PROVIDENCE ASP-34.  
 RX MEDLINE=87175640; PubMed=3470784;  
 RA Chan S.J., Seino S., Gruppuso P.A., Schwartz R., Steiner D.F.;  
 RT "A mutation in the B chain coding region is associated with impaired  
 RT proinsulin conversion in a family with hyperproinsulinemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).  
 RN [18]  
 RP VARIANT WAKAYAMA LEU-92.  
 RX MEDLINE=87058122; PubMed=3537011;  
 RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;  
 RT "Structurally abnormal insulin in a diabetic patient. Characterization  
 RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";  
 RL J. Clin. Invest. 78:1666-1672(1986).  
 RN [19]  
 RP VARIANT HIS-89.  
 RX MEDLINE=90317021; PubMed=2196279;  
 RA Barbetti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,  
 RA Merenich J.A., Taylor S.I., Roth J.;  
 RT "Two unrelated patients with familial hyperproinsulinemia due to a  
 RT mutation substituting histidine for arginine at position 65 in the  
 RT proinsulin molecule: identification of the mutation by direct  
 RT sequencing of genomic deoxyribonucleic acid amplified by polymerase  
 RT chain reaction.";  
 RL J. Clin. Endocrinol. Metab. 71:164-169(1990).  
 RN [20]  
 RP VARIANT HIS-89.  
 RX MEDLINE=85261996; PubMed=4019786;  
 RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;  
 RT "Posttranslational cleavage of proinsulin is blocked by a point  
 RT mutation in familial hyperproinsulinemia.";  
 RL J. Clin. Invest. 76:378-380(1985).  
 RN [21]  
 RP VARIANT KYOTO LEU-89.  
 RX MEDLINE=92291307; PubMed=1601997;  
 RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;  
 RT "A novel point mutation in the human insulin gene giving rise to  
 RT hyperproinsulinemia (proinsulin Kyoto).";  
 RL J. Clin. Invest. 89:1902-1907(1992).  
 RN [22]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=91104966; PubMed=2271664;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Toward the solution structure of human insulin: sequential 2D 1H NMR  
 RT assignment of a des-pentapeptide analogue and comparison with crystal  
 RT structure.";  
 RL Biochemistry 29:10545-10555(1990).  
 RN [23]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=91242467; PubMed=2036420;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Comparative 2D NMR studies of human insulin and des-pentapeptide  
 RT insulin: sequential resonance assignment and implications for protein  
 RT dynamics and receptor recognition.";  
 RL Biochemistry 30:5505-5515(1991).

RN [24]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=91265527; PubMed=1646635;  
 RA Hua Q.-X., Weiss M.A.;  
 RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-  
 RT specific resonance assignments and effects of solvent composition.";  
 RL Biochim. Biophys. Acta 1078:101-110(1991).

Query Match 100.0%; Score 463; DB 1; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-42;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |  
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

## RESULT 2

INS\_PANTR

ID INS\_PANTR STANDARD; PRT; 110 AA.

AC P30410;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Insulin precursor.

GN INS.

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI\_TaxID=9598;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92219953; PubMed=1560757;

RA Seino S., Bell G.I., Li W.;

RT "Sequences of primate insulin genes support the hypothesis of a  
 RT slower rate of molecular evolution in humans and apes than in  
 RT monkeys.";

RL Mol. Biol. Evol. 9:193-203(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=22833521; PubMed=12952878;

RA Stead J.D., Hurles M.E., Jeffreys A.J.;

RT "Global haplotype diversity in the human insulin gene region.";  
 RL Genome Res. 13:2101-2111(2003).

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

CC -----



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; X61089; CAA43403.1; -.  
 DR EMBL; AY137497; AAN06933.1; -.  
 DR PIR; A42179; A42179.  
 DR PDB; 1EFE; 29-MAR-00.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEF5 CRC64;

Query Match 100.0%; Score 463; DB 1; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-42;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||||||||||||||||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 ||||||||||||||||  
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

# RESULT 3

INS\_CERAE

ID INS\_CERAE STANDARD; PRT; 110 AA.

AC P30407; P01309;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS Cercopithecus aethiops (Green monkey) (Grivet).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

OC Cercopithecinae; Cercopithecus.

OX NCBI\_TaxID=9534;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92219953; PubMed=1560757;



Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 4

INS\_MACFA

ID INS\_MACFA STANDARD; PRT; 110 AA.

AC P30406; P01309;

DT 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

OC Cercopithecinae; Macaca.

OX NCBI\_TaxID=9541;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=83080474; PubMed=6184262;

RA Wetekam W., Groneberg J., Leineweber M., Wengenmayer F.,

RA Winnacker E.-L.;

RT "The nucleotide sequence of cDNA coding for preproinsulin from the  
RT primate Macaca fascicularis.";

RL Gene 19:179-183(1982).

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
CC increases cell permeability to monosaccharides, amino acids and  
CC fatty acids. It accelerates glycolysis, the pentose phosphate  
CC cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
CC disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; J00336; AAA36849.1; -.

DR PIR; JQ0178; JQ0178.

DR HSSP; P01308; 1AI0.

DR InterPro; IPR004825; Ins/IGF/relax.

DR Pfam; PF00049; Insulin; 1.

DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;

Query Match 98.5%; Score 456; DB 1; Length 110;  
 Best Local Similarity 98.8%; Pred. No. 6.4e-42;  
 Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEG 84  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |||  
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

# RESULT 5

## INS\_RABIT

ID INS\_RABIT STANDARD; PRT; 110 AA.  
 AC P01311;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Pancreas;  
 RX MEDLINE=94179230; PubMed=8132571;  
 RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,  
 RA Menon R.K., Zahm D.S.;  
 RT "Insulin gene expression and insulin synthesis in mammalian neuronal  
 RT cells.";  
 RL J. Biol. Chem. 269:8445-8454(1994).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE=66160119; PubMed=5949593;  
 RA Smith L.F.;  
 RT "Species variation in the amino acid sequence of insulin.";  
 RL Am. J. Med. 40:662-666(1966).  
 RN [3]  
 RP SEQUENCE OF 56-110 FROM N.A.  
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;  
 RL Submitted (APR-1991) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.





## RESULT 7

## INS\_SPETR

ID INS\_SPETR STANDARD; PRT; 110 AA.

AC Q9LXI3;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS *Spermophilus tridecemlineatus* (Thirteen-lined ground squirrel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;

OC *Spermophilus*.

OX NCBI\_TaxID=43179;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RA Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;

RT "Regulation of PDK4 expression in a hibernating mammal.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AY038604; AAK72558.1; -.

DR HSSP; P01308; 1LNP.

DR InterPro; IPR004825; Ins/IGF/relax.

DR Pfam; PF00049; Insulin; 1.

DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 24 BY SIMILARITY.

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).

FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).

FT DISULFID 95 100 BY SIMILARITY.

SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 89.2%; Score 413; DB 1; Length 110;  
 Best Local Similarity 89.5%; Pred. No. 2.5e-37;  
 Matches 77; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |||||:| | |||||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSRREVEEQQGGQVELGGGPGAGLPQPLALEM 84

QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 :|||  
 Db 85 ALQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 8

INS\_HORSE

ID INS\_HORSE STANDARD; PRT; 86 AA.  
 AC P01310;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE OF 1-30 AND 66-86.  
 RA Harris J.I., Sanger F., Naughton M.A.;  
 RT "Species differences in insulin.";  
 RL Arch. Biochem. Biophys. 65:427-438(1956).  
 RN [2]  
 RP SEQUENCE OF 33-63.  
 RX MEDLINE=73061498; PubMed=4640931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 RT and the horse.";  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -!- CAUTION: X'S AT POSITIONS 31-32 AND 64-65 REPRESENT PAIRED BASIC  
 CC RESIDUES ASSUMED BY HOMOLOGY TO BE PRESENT IN THE PRECURSOR  
 CC MOLECULE.  
 DR PIR; A01580; IPHO.  
 DR HSSP; P01317; IAPH.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism.



FT	CHAIN	1	30	INSULIN B CHAIN.
FT	PROPEP	33	63	C PEPTIDE.
FT	CHAIN	66	86	INSULIN A CHAIN.
FT	DISULFID	7	72	INTERCHAIN.
FT	DISULFID	19	85	INTERCHAIN.
FT	DISULFID	71	76	
SQ	SEQUENCE	86 AA;	9142 MW;	A3E1E822711BDB46 CRC64;

Query Match 85.1%; Score 394; DB 1; Length 86;  
Best Local Similarity 84.9%; Pred. No. 2.1e-35;  
Matches 73; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGQVELGGGPGAGSLQPLALEG 60
        |||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKAXXEAEDPQVGEVELGGGPGGLQPLALAG 60
        |||||

Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
        | |||||
Db      61 POOXXGIVEOCCTGICSLYQLENYCN 86

```

## RESULT 9

```

INS2_MOUSE
ID INS2_MOUSE STANDARD; PRT; 110 AA.
AC P01326;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin 2 precursor.
GN INS2 OR INS-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87169768; PubMed=3104603;
RA Wentworth B.M., Schaefer I.M., Villa-Komaroff L., Chirgwin J.M.;
RT "Characterization of the two nonallelic genes encoding mouse
RT preproinsulin.";
RL J. Mol. Evol. 23:305-312(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NON;
RX MEDLINE=90372989; PubMed=2397023;
RA Sawa T., Ohgaku S., Morioka H., Yano S.;
RT "Molecular cloning and DNA sequence analysis of preproinsulin genes
RT in the NON mouse, an animal model of human non-obese, non-insulin-
RT dependent diabetes mellitus.";
RL J. Mol. Endocrinol. 5:61-67(1990).
RN [3]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=72189455; PubMed=5063718;
RA Buenzli H.F., Glatthaar B., Kunz P., Muelhaupt E., Humbel R.E.;
RT "Amino acid sequence of the two insulins from mouse (Maus musculus).";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458(1972).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It

```

CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04724; CAA28433.1; -.  
 DR PIR; A26342; INMS2.  
 DR HSSP; P01317; 1APH.  
 DR MGD; MGI:96573; Ins2.  
 DR GO; GO:0000187; P:activation of MAPK; IDA.  
 DR GO; GO:0042325; P:regulation of phosphorylation; IDA.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 2 B CHAIN.  
 FT PROPEP 57 87 INSULIN 2 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 2 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12364 MW; 3554C8803D24FDAD CRC64;

Query Match 85.1%; Score 394; DB 1; Length 110;  
 Best Local Similarity 84.9%; Pred. No. 2.7e-35;  
 Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 || ||||| :||| || || :||| || || ||  
 Db 25 FVKQHLCGSHLVEALYLVCGERGFFYTPMSRREVDPQVAQLLEGGGPGAGDLQTLALEV 84  
 QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 : ||||| :||| ||||| |||||  
 Db 85 AQQKRGIVDQCCTSICSLYQLENYCN 110

# RESULT 10

INS2\_RAT

ID INS2\_RAT STANDARD; PRT; 110 AA.  
 AC P01323;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin 2 precursor.  
 GN INS2 OR INS-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=80045035; PubMed=498284;  
 RA Lomedico P., Rosenthal N., Efstratiadis A., Gilbert W., Kolodner R.,  
 RA Tizard R.;  
 RT "The structure and evolution of the two nonallelic rat preproinsulin  
 RT genes.";  
 RL Cell 18:545-558(1979).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86310882; PubMed=2427930;  
 RA Soares M.B., Schin E., Henderson A., Karathanasis S.K., Cate R.,  
 RA Zeitlin S., Chirgwin J., Efstratiadis A.;  
 RT "RNA-mediated gene duplication: the rat preproinsulin I gene is a  
 RT functional retroposon.";  
 RL Mol. Cell. Biol. 5:2090-2103(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=80240379; PubMed=6249167;  
 RA Lomedico P.T., Rosenthal N., Kolodner R., Efstratiadis A.,  
 RA Gilbert W.;  
 RT "The structure of rat preproinsulin genes.";  
 RL Ann. N.Y. Acad. Sci. 343:425-432(1980).  
 RN [4]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE=70067613; PubMed=4311938;  
 RA Steiner D.F., Clark J.L., Nolan C., Rubenstein A.H., Margoliash E.,  
 RA Aten B., Oyer P.E.;  
 RT "Proinsulin and the biosynthesis of insulin.";  
 RL Recent Prog. Horm. Res. 25:207-282(1969).  
 RN [5]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE=73061498; PubMed=4640931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 RT and the horse.";  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 RN [6]  
 RP SEQUENCE OF 57-87, AND REVISIONS.  
 RX MEDLINE=72177385; PubMed=4554104;  
 RA Markussen J., Sundby F.;  
 RT "Rat-proinsulin C-peptides. Amino-acid sequences.";  
 RL Eur. J. Biochem. 25:153-162(1972).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; V01243; CAA24560.1; -.  
 DR EMBL; J00748; AAA41443.1; -.  
 DR EMBL; M25585; AAA41440.1; -.  
 DR EMBL; M25583; AAA41440.1; JOINED.  
 DR PIR; B90789; IPRT2.  
 DR HSSP; P01317; IAPH.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 2 B CHAIN.  
 FT PROPEP 57 87 INSULIN 2 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 2 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100  
 SQ SEQUENCE 110 AA; 12339 MW; 3A626DA98C86F3CA CRC64;

Query Match 85.1%; Score 394; DB 1; Length 110;  
 Best Local Similarity 84.9%; Pred. No. 2.7e-35;  
 Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 || ||||| :|| || || :||| || || || ||  
 Db 25 FVKQHLCGSHLVEALYLVCGERGFFYTPMSRREVEDPQVAQLELGGGPGAGDLQTLALEV 84  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 : |||||:|||||  
 Db 85 ARQKRGIVDQCCTSICSLYQLENYCN 110

# RESULT 11

INS\_AOTTR

ID INS\_AOTTR STANDARD; PRT; 108 AA.

AC P10604;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS Aotus trivirgatus (Night monkey) (Douroucoulis), and

OS Saimiri sciureus (Common squirrel monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.

OX NCBI\_TaxID=9505, 9521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=A.trivirgatus;  
 RX MEDLINE=88041119; PubMed=3118367;  
 RA Seino S., Steiner D.F., Bell G.I.;  
 RT "Sequence of a New World primate insulin having low biological  
 RT potency and immunoreactivity."  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7423-7427(1987).  
 RN [2]  
 RP SEQUENCE OF 25-54 AND 88-108.  
 RC SPECIES=S.sciureus;  
 RX MEDLINE=91088593; PubMed=2263627;  
 RA Yu J.-H., Eng J., Yalow R.S.;  
 RT "Isolation and amino acid sequences of squirrel monkey (Saimiri  
 RT sciurea) insulin and glucagon."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9766-9768(1990).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J02989; AAA35374.1; -.  
 DR PIR; A39883; A39883.  
 DR HSSP; P01308; 1HIS.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 85 C PEPTIDE.  
 FT CHAIN 88 108 INSULIN A CHAIN.  
 FT DISULFID 31 94 INTERCHAIN.  
 FT DISULFID 43 107 INTERCHAIN.  
 FT DISULFID 93 98  
 SQ SEQUENCE 108 AA; 11842 MW; 1869B8250099731F CRC64;

Query Match 84.7%; Score 392; DB 1; Length 108;  
 Best Local Similarity 84.9%; Pred. No. 4.3e-35;  
 Matches 73; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60



KW	Insulin family; Hormone; Glucose metabolism; Signal.		
FT	SIGNAL	1	24
FT	CHAIN	25	54 INSULIN B CHAIN.
FT	PROPEP	57	87 C PEPTIDE.
FT	CHAIN	90	110 INSULIN A CHAIN.
FT	DISULFID	31	96 INTERCHAIN.
FT	DISULFID	43	109 INTERCHAIN.
FT	DISULFID	95	100
SQ	SEQUENCE 110 AA; 12268 MW; 219E92B85A535CEC CRC64;		

Query Match 84.7%; Score 392; DB 1; Length 110;  
Best Local Similarity 84.9%; Pred. No. 4.4e-35;  
Matches 73; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

```

Qy      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAE DLQVGQVELGGGPGAGSLQPLALEG 60
      |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSRRGVEDPQVAQLELGGGPGADDLQTLALEV 84
      |||
Qy      61 SLQKRGIVEQCCTSICSLYQLENYCN 86
      : |||
Db      85 AQQKRGIVDQCCTSICSLYQLENYCN 110

```

RESULT 13

```

INS1_RAT
ID      INS1_RAT          STANDARD;          PRT;    110 AA.
AC      P01322;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Insulin 1 precursor.
GN      INS1 OR INS-1.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=80045034; PubMed=498283;
RA      Cordell B., Bell G.I., Tischer E., Denoto F.M., Ullrich A.,
RA      Pictet R.L., Rutter W.J., Goodman H.M.;
RT      "Isolation and characterization of a cloned rat insulin gene.";
RL      Cell 18:533-543(1979).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Liver;
RX      MEDLINE=80045035; PubMed=498284;
RA      Lomedico P., Rosenthal N., Efstratiadis A., Gilbert W., Kolodner R.,
RA      Tizard R.;
RT      "The structure and evolution of the two nonallelic rat preproinsulin
RT      genes.";
RL      Cell 18:545-558(1979).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=80240379; PubMed=6249167;
RA      Lomedico P.T., Rosenthal N., Kolodner R., Efstratiadis A.,
RA      Gilbert W.;

```

RT "The structure of rat preproinsulin genes.";  
 RL Ann. N.Y. Acad. Sci. 343:425-432(1980).  
 RN [4]  
 RP SEQUENCE OF 25-54 AND 90-110.  
 RX MEDLINE=70067613; PubMed=4311938;  
 RA Steiner D.F., Clark J.L., Nolan C., Rubenstein A.H., Margoliash E.,  
 RA Aten B., Oyer P.E.;  
 RT "Proinsulin and the biosynthesis of insulin.";  
 RL Recent Prog. Horm. Res. 25:207-282(1969).  
 RN [5]  
 RP SEQUENCE OF 57-87.  
 RX MEDLINE=73061498; PubMed=4640931;  
 RA Tager H.S., Steiner D.F.;  
 RT "Primary structures of the proinsulin connecting peptides of the rat  
 RT and the horse.";  
 RL J. Biol. Chem. 247:7936-7940(1972).  
 RN [6]

RP SEQUENCE OF 57-87, AND REVISIONS.  
 RX MEDLINE=72177385; PubMed=4554104;  
 RA Markussen J., Sundby F.;  
 RT "Rat-proinsulin C-peptides. Amino-acid sequences.";  
 RL Eur. J. Biochem. 25:153-162(1972).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; V01242; CAA24559.1; -.  
 DR EMBL; J00747; AAA41442.1; -.  
 DR EMBL; M25584; AAA41439.1; -.  
 DR PIR; A90788; IPRT1.  
 DR HSSP; P01308; 1A7F.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN 1 B CHAIN.  
 FT PROPEP 57 87 INSULIN 1 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN 1 A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN.  
 FT DISULFID 43 109 INTERCHAIN.  
 FT DISULFID 95 100



SQ SEQUENCE 110 AA; 12420 MW; 51D606DA54AE3533 CRC64;

Query Match 83.2%; Score 385; DB 1; Length 110;  
Best Local Similarity 83.7%; Pred. No. 2.4e-34;  
Matches 72; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
|| ||||| |||||||||||||||||||:|||| || || |:||||| || || |||||  
Db 25 FVKQHLCGPHLVEALYLVCGERGFFYTPKSRREVEDPQVPQLELGGGPEAGDLQTLALEV 84  
  
QY 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
: |||||||:|||||||||||||||||  
Db 85 ARQKRGIVDQCCTSICSLYQLENYCN 110

#### RESULT 14

##### INS\_PIG

ID INS\_PIG STANDARD; PRT; 108 AA.  
AC P01315; Q9TSJ5;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Insulin precursor.  
GN INS.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Han X.G., Tuch B.E.;  
RT "Complete porcine preproinsulin cDNA sequence.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Large white;  
RX MEDLINE=22135958; PubMed=12140686;  
RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,  
RA Georges M., Andersson L.;  
RT "Comparative sequence analysis of the INS-IGF2-H19 gene cluster in  
RT pigs.";  
RL Mamm. Genome 13:388-398(2002).  
RN [3]  
RP SEQUENCE OF 25-108.  
RX MEDLINE=68286485; PubMed=5657063;  
RA Chance R.E., Ellis R.M., Bromer W.W.;  
RT "Porcine proinsulin: characterization and amino acid sequence.";  
RL Science 161:165-167(1968).  
RN [4]  
RP REVISION TO 59.  
RA Chance R.E.;  
RL Submitted (JUL-1970) to the PIR data bank.  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;  
RT "Insulin. The structure in the crystal and its reflection in  
RT chemistry and biology.";

RL Adv. Protein Chem. 26:279-402(1972).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Isaacs N.W., Agarwal R.C.;  
 RT "Experience with fast Fourier least squares in the refinement of the  
 RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A  
 RT resolution.";  
 RL Acta Crystallogr. A 34:782-791(1978).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RX MEDLINE=89099318; PubMed=2905485;  
 RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,  
 RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,  
 RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;  
 RT "The structure of 2Zn pig insulin crystals at 1.5-A resolution.";  
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=92126280; PubMed=1772633;  
 RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;  
 RT "Structure of porcine insulin cocrystallized with clupeine Z.";  
 RL Acta Crystallogr. B 47:975-986(1991).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE=91222450; PubMed=2025410;  
 RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,  
 RA Dodson G.G., North A.C.T.;  
 RT "Structure of the pig insulin dimer in the cubic crystal.";  
 RL Acta Crystallogr. B 47:127-136(1991).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
 RA Diao J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;  
 RT "Structure of monomeric porcine DesB1-B2 despentapeptide (B26-B30)  
 RT insulin at 1.65-A resolution.";  
 RL Acta Crystallogr. D 53:507-512(1997).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -!- DATABASE: NAME=Protein Spotlight;  
 CC NOTE=Issue 9 of April 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/sptlt009.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF064555; AAC77920.1; ALT\_INIT.  
 DR EMBL; AY044828; AAL69550.1; -.

DR PDB; 3INS; 09-JAN-89.  
 DR PDB; 4INS; 31-JUL-94.  
 DR PDB; 6INS; 31-JAN-94.  
 DR PDB; 7INS; 31-JAN-94.  
 DR PDB; 9INS; 15-OCT-91.  
 DR PDB; 1IZA; 15-OCT-91.  
 DR PDB; 1IZB; 15-OCT-91.  
 DR PDB; 2TCI; 29-JAN-96.  
 DR PDB; 1MPJ; 29-JAN-96.  
 DR PDB; 3MTH; 29-JAN-96.  
 DR PDB; 1DEI; 16-JUN-97.  
 DR PDB; 1SDB; 01-APR-98.  
 DR PDB; 1WAV; 28-FEB-97.  
 DR PDB; 1ZEI; 16-FEB-99.  
 DR PDB; 1ZNI; 28-JAN-98.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 85 C PEPTIDE.  
 FT CHAIN 88 108 INSULIN A CHAIN.  
 FT DISULFID 31 94 INTERCHAIN.  
 FT DISULFID 43 107 INTERCHAIN.  
 FT DISULFID 93 98  
 FT HELIX 26 46  
 FT STRAND 48 48  
 FT HELIX 89 94  
 FT HELIX 100 106  
 FT STRAND 107 107  
 SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBE CRC64;

Query Match 82.7%; Score 383; DB 1; Length 108;  
 Best Local Similarity 86.0%; Pred. No. 3.9e-34;  
 Matches 74; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 |||||||||||||||||||||||||||||: || ||||| || || |||||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGG--GLGGLQALALEG 82  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86  
 |||||||||||||||||||||  
 Db 83 PPQKRGIVEQCCTSICSLYQLENYCN 108

RESULT 15  
 INS\_PSAOB  
 ID INS\_PSAOB STANDARD; PRT; 110 AA.  
 AC Q62587;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Insulin precursor.  
 GN INS.

OS Psammomys obesus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
 OC Psammomys.  
 OX NCBI\_TaxID=48139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=97309250; PubMed=9166665;  
 RA Kaiser N., Baillyes E.M., Schneider B.S., Cerasi E., Steiner D.F.,  
 RA Hutton J.C., Gross D.J.;  
 RT "Characterization of the unusual insulin of Psammomys obesus, a  
 RT rodent with nutrition-induced NIDDM-like syndrome."  
 RL Diabetes 46:953-957(1997).  
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the insulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X98241; CAA66897.1; -.  
 DR HSSP; P01308; 1AI0.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR Pfam; PF00049; Insulin; 1.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KW Insulin family; Hormone; Glucose metabolism; Signal.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 54 INSULIN B CHAIN.  
 FT PROPEP 57 87 C PEPTIDE.  
 FT CHAIN 90 110 INSULIN A CHAIN.  
 FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 95 100 BY SIMILARITY.  
 SQ SEQUENCE 110 AA; 12324 MW; A006738E20579CB0 CRC64;

Query Match 81.4%; Score 377; DB 1; Length 110;  
 Best Local Similarity 81.4%; Pred. No. 1.7e-33;  
 Matches 70; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 60  
 ||||| :| |:| :||| ||||| :|||  
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKFRRGVDDPQMPQLELGGSPGAGDLRALALEV 84  
 Qy 61 SLQKRGIVEQCCTSICSLYQLENYCN 86

Db                   : |||||       |||  
85 ARQKRGIVEQCCTGICSLYQLENYCN 110

Search completed: July 15, 2004, 16:36:25  
Job time : 5.93657 secs